GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 00:55:20 ; Search time 3253 Seconds
(without alignments)
15871.004 Million cell updates/sec

Title: US-09-854-300-7
Perfect score: 1774
Sequence: 1 agctggagctccaccgcggt.....ataggcaagtttcctctcag 1774

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1662.8	93.7	2796	6	AX337968	AX337968 Sequence
2		1659.6	93.6	2830	9	AK074264	AK074264 Homo sapi
3		1280.8	72.2	1297	9	AF394689	AF394689 Homo sapi
4		1205.2	67.9	1928	10	BC010477	BC010477 Mus muscu
5		1203.6	67.8	1936	10	AB041548	AB041548 Mus muscu
6		1106.8	62.4	1287	10	AF426411	AF426411 Mus muscu
7		714.4	40.3	1871	9	AK027169	AK027169 Homo sapi
8		674.4	38.0	169812	9	AL391315	AL391315 Human DNA
9		674.4	38.0	169863	2	AL591467	AL591467 Homo sapi
10		392.2	22.1	404	6	AX071980	AX071980 Sequence
11		372.6	21.0	646	6	AX341568	AX341568 Sequence
12	c	310	17.5	435	6	AX396961	AX396961 Sequence
13	c	280.4	15.8	195263	2	AC115936	AC115936 Mus muscu
14	c	278.4	15.7	167021	2	AC111461	AC111461 Rattus no
15		268.8	15.2	1391	9	AB070023	AB070023 Macaca fa
16		268.8	15.2	1395	9	AB070058	AB070058 Macaca fa
17		266.8	15.0	1429	9	AK098524	AK098524 Homo sapi
18		266.8	15.0	1430	9	BC022038	BC022038 Homo sapi
19		266.8	15.0	1446	9	AF447589	AF447589 Homo sapi
20	c	266.8	15.0	120984	9	AC006463	AC006463 Homo sapi
21		256.4	14.5	1328	9	BC029264	BC029264 Homo sapi
22		254.8	14.4	1355	6	AX061659	AX061659 Sequence
23		253.2	14.3	1261	9	AK098654	AK098654 Homo sapi
24	c	220.2	12.4	116739	2	AC114183	AC114183 Rattus no
25	c	220.2	12.4	140346	2	AC109686	AC109686 Rattus no
26		220	12.4	119995	10	AL731648	AL731648 Mouse DNA
27		217.4	12.3	116739	2	AC114183	AC114183 Rattus no
28		209	11.8	1200	6	AX083416	AX083416 Sequence
29		209	11.8	1527	6	AX083426	AX083426 Sequence
30		209	11.8	2001	9	BC019355	BC019355 Homo sapi
31		207.4	11.7	2013	9	BC032328	BC032328 Homo sapi
32		156.8	8.8	839	9	AK058116	AK058116 Homo sapi
33		148.8	8.4	1452	10	AF171875	AF171875 Mus muscu
34		148.8	8.4	1506	10	BC018199	BC018199 Mus muscu
35		147.6	8.3	1153	9	AF155650	AF155650 Human mRN
36		147.6	8.3	1253	6	AR030960	AR030960 Sequence
37		147.6	8.3	1257	6	AX083320	AX083320 Sequence
38		147.6	8.3	1433	6	AX083330	AX083330 Sequence
39		147.6	8.3	1445	9	BC017100	BC017100 Homo sapi
40	c	140	7.9	118766	2	AC128696	AC128696 Rattus no
41		108.4	6.1	5111	9	AB033040	AB033040 Homo sapi
42		105.6	6.0	382	6	AX261677	AX261677 Sequence
43		87.2	4.9	2369	3	DROGOLTHA	M97204 Fruitfly go
44		84.8	4.8	2586	3	AY069169	AY069169 Drosophil
45		79	4.5	2448	4	OCU132110	AJ132110 Oryctolag

ALIGNMENTS

RESULT 1
AX337968
LOCUS AX337968 2796 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 9 from Patent WO0194391.
ACCESSION AX337968
VERSION AX337968.1 GI:18128679
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yue,H., He,A., Nguyen,D.B., Yao,M.G., Bandman,O., Burford,N.,
Tang,Y.T., Xu,Y., Hafalia,A., Azimzai,Y. and Walia,N.K.
TITLE Intracellular signaling proteins

```
  JOURNAL   Patent: WO 0194391-A 9 13-DEC-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2796
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2791668CB1"
BASE COUNT       821 a    548 c    620 g    807 t
ORIGIN

  Query Match             93.7%;  Score 1662.8;  DB 6;  Length 2796;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1690;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    2;

Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 87

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 147

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 207

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     208 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 267

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     268 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 327

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     328 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 387

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     388 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 447

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     448 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 507

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     508 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 567

Qy     614 GCCCTCATCCAACGCGGCGGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     568 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 627

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     628 AGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 687

Qy     734 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     688 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 747

Qy     794 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     748 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 807

Qy     854 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     808 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 867

Qy     914 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     868 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 927

Qy     974 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     928 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 987

Qy    1034 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     988 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1047

Qy    1094 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1048 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1107

Qy    1154 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGC 1213
           |||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db     1108 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTGC 1167

Qy    1214 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1168 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1227

Qy    1274 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1228 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1287

Qy    1334 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1288 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1347

Qy    1394 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1348 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1407

Qy    1454 GTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1513
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     1408 GTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1467

Qy    1514 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1468 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1527

Qy    1574 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1528 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1587

Qy    1634 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1588 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1647

Qy    1694 AATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATATTTTT 1753
           |   | | ||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     1648 A--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATAATATTTTT 1704

Qy    1754 CATAGGCAAGTTTCCTCTCAG 1774
           |||||||||||||||||||||
Db     1705 CATAGGCAAGTTTCCTCTCAG 1725


RESULT 2
AK074264
LOCUS       AK074264               2830 bp    mRNA    linear   PRI 15-FEB-2002
DEFINITION  Homo sapiens cDNA FLJ23684 fis, clone HEP09821.
ACCESSION   AK074264
VERSION     AK074264.1  GI:18676818
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
            clone:HEP09821.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
```

```
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2830)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2002) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure, Human Genome
            Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology; cDNA library construction,
            5'- & 3'-end one pass sequencing: Departent of Virology and Human
            Genome Center, Institute of Medical Science, University of Tokyo
            (partly supported by Science and Technology Agency).
FEATURES             Location/Qualifiers
     source          1. .2830
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEP09821"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone_lib="HEP"
                     /note="cloning vector: pME18SFL3"
     CDS             251. .1537
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB85033.1"
                     /db_xref="GI:18676819"
                     /translation="MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWT
                     AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNF
                     TVPTVWGSTVQVSWLALIQRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSH
                     PGAVDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFPVSVSFFII
                     TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDS
                     CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG
                     SVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVN
                     HEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS"
BASE COUNT      827 a    564 c    630 g    809 t
ORIGIN

  Query Match           93.6%;  Score 1659.6;  DB 9;  Length 2830;
  Best Local Similarity  99.2%;  Pred. No. 0;
  Matches 1688;  Conservative    0;  Mismatches   10;  Indels    3;  Gaps    2;

Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db      62 CCGAGGAGCTGCATCTGCGGCAACCTGTGCGCTGACGCTACGTGCCTCCTGGCTCCGACG 121

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 181

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 241

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 301

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 361

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 421

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 481

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 541

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 601

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 661

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 721

Qy     734 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 781

Qy     794 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 841

Qy     854 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 901

Qy     914 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     902 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 961

Qy     974 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     962 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 1021

Qy    1034 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1022 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1081

Qy    1094 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1082 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1141

Qy    1154 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGC 1213
           |||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||
Db    1142 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTGC 1201

Qy    1214 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1202 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1261

Qy    1274 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1262 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1321

Qy    1334 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1393
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    1322 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACGGATGAACCGCCTCTG 1381

Qy    1394 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1382 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1441

Qy    1454 GTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1513
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    1442 GTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1501

Qy    1514 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1502 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1561
```

```
Qy   1574 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1633
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1562 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1621

Qy   1634 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1622 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1681

Qy   1694 AATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATATTTTT 1753
          |   |  | ||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1682 A--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATAATATTTTT 1738

Qy   1754 CATAGGCAAGTTTCCTCTCAG 1774
          ||||||||||||||||||||
Db   1739 CATAGGCAAGTTTCCTCTCAG 1759
```

```
RESULT 3
AF394689
LOCUS       AF394689               1297 bp    mRNA    linear   PRI 25-JUL-2001
DEFINITION  Homo sapiens GRAIL (GRAIL) mRNA, complete cds.
ACCESSION   AF394689
VERSION     AF394689.1  GI:15011451
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1297)
  AUTHORS   Soares,L.R. and Fathman,G.
  TITLE     Grail: a novel ring finger protein upregulated in anergic T cells
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1297)
  AUTHORS   Soares,L.R. and Fathman,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2001) Pathology, Stanford University, 800 Welch
            Rd, Palo Alto, CA 94304, USA
FEATURES             Location/Qualifiers
     source          1. .1297
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
     gene            1. .1297
                     /gene="GRAIL"
     CDS             11. .1297
                     /gene="GRAIL"
                     /note="E3, ring finger ubiquitin ligase; regulated in
                     anergy-induced lymphocytes"
                     /codon_start=1
                     /product="GRAIL"
                     /protein_id="AAK77554.1"
                     /db_xref="GI:15011452"
                     /translation="MGPLPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWT
                     AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNF
                     TVPTVWGSTVQVSWLALIQRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSH
                     PGAVDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVPFFII
                     TXATGGYFIFYSAXRLRNARXQSXKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDS
                     CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG
                     SVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVN
                     HEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS"
BASE COUNT      324 a    311 c    355 g    303 t      4 others
ORIGIN

  Query Match           72.2%;  Score 1280.8;  DB 9;  Length 1297;
  Best Local Similarity  99.1%;  Pred. No. 5.8e-283;
  Matches 1285;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy    253 CGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTT 312
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      1 CGCGCGCGCCATGGGGCCGCTGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTT 60

Qy    313 TTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCG 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCG 120

Qy    373 GGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGG 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGG 180

Qy    433 AGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCT 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCT 240

Qy    493 GGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAA 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGAGGGCCCGGGGCGCTTAACGCCTGTAA 300

Qy    553 CCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTT 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTT 360

Qy    613 GGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGA 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGA 420

Qy    673 GAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCC 732
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCC 480

Qy    733 CATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCAC 792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCAC 540

Qy    793 AAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAA 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAA 600

Qy    853 AAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTAT 912
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    601 AAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGCCCTTTTTTAT 660

Qy    913 TATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGC 972
          |||||||| |||||||| ||||||||||||||||||||||||||||| ||||||||||||
Db    661 TATTACGGNGGCAACTGGGGGCTATTTTATCTTTTATTCTGCTCNAAGGCTACGGAATGC 720

Qy    973 AAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAG 1032
          |||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AAGANCTCAAAGCNGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAG 780

Qy   1033 GCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTG 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTG 840

Qy   1093 TGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCA 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCA 900

Qy   1153 TATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTG 1212
          |||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db    901 TATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTG 960

Qy   1213 CAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTC 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTC 1020

Qy   1273 TTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAA 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAA 1080

Qy   1333 TCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCT 1392
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   1081 TCGCAGCGAGACCGCATCATCTGGATACGCTTCAGTACAGGGAACAGATGAACCGCCTCT 1140

Qy   1393 GGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTC 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1141 GGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTC 1200

Qy    1453 TGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCC 1512
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    1201 TGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCC 1260

Qy    1513 TAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAA 1549
           |||||||||||||||||||||||||||||||||||||
Db    1261 TAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAA 1297


RESULT 4
BC010477
LOCUS       BC010477                1928 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, RIKEN cDNA 1300002C13 gene, clone MGC:19395
            IMAGE:3154433, mRNA, complete cds.
ACCESSION   BC010477
VERSION     BC010477.1  GI:14714672
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1928)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 23 Row: b Column: 12
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 12963610.
FEATURES             Location/Qualifiers
     source          1. .1928
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="129,C57BL/6J,FVB/N"
                     /clone="MGC:19395 IMAGE:3154433"
                     /tissue_type="Mammary tumor. Brca1-/fl; MMTV-Cre model. 10
                     months old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam3"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             176. .1462
                     /codon_start=1
                     /product="RIKEN cDNA 1300002C13 gene"
                     /protein_id="AAH10477.1"
                     /db_xref="GI:14714673"
                     /db_xref="LocusID:66889"
                     /translation="MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWT
                     AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNF
                     TVPTVWGSTVQVSWLALIQRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSH
                     PGAGDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFII
                     TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDS
                     CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG
                     SVSLQVPVSNEASNTASPHEEDSRSETASSGYASVQGADEPPLEEHAQSANENLQLVN
                     HEANSVAVDVVPHVDNPTFEEDETPDQEAAVREIKS"
BASE COUNT      482 a    466 c    498 g    482 t
ORIGIN

  Query Match           67.9%;  Score 1205.2;  DB 10;  Length 1928;
  Best Local Similarity  89.6%;  Pred. No. 1.4e-265;
  Matches 1364;  Conservative    0;  Mismatches  149;  Indels   10;  Gaps    6;

Qy     252 GCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCT 311
           | | ||||||||||||||||||||| ||| |||||||| ||||||||| ||||||||
Db     165 GAGTGCGCGCCATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAG 224

Qy     312 TTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCC 371
           | || |  | ||||| |||||||||||| |||| |||||||||||| || ||||||||||
Db     225 CTGCCCGGCTACTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCC 284

Qy     372 GGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGG 431
           |||| || ||||| ||||||||||||||||||||||||||||||||||||||||||||| |
Db     285 GGGGAGCCGAAGCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCG 344

Qy     432 GAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 491
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     345 GAGTGAACCGCACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 404

Qy     492 TGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTA 551
           ||||||| ||  | ||||||||||||||||||||||||||||||||||| ||||||||||
Db     405 TGGAGCCCGTCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTA 464

Qy     552 ACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGT 611
           |||||||||| |||||||||||||||||||||||||||| |||| |||||||| |||||||
Db     465 ACCCGCACACCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGT 524

Qy     612 TGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATG 671
           |||||||||||||||||| || |||||||||||||| ||||||||||||||||||||  |
Db     525 TGGCCCTCATCCAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAG 584

Qy     672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
           ||||| |||| |||||||| |||||||||||||||| ||||||||||||||||||||||
Db     585 AGAGAGGGGCTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCC 644

Qy     732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
           |||||||||||||||||||| | ||||||||||||||||||||| |||||||||||||||
Db     645 CCATGTCTCACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAA 704

Qy     792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
           |||||||||||| |||||||||||||||| |||| |||||||||| ||||||||||||||
Db     705 CAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGA 764

Qy     852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTA 911
           |||||||||||||||||||||| |||||||||| |||||||||||||||||||||||| |
Db     765 AAAAACATGGCCCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCA 824

Qy     912 TTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATG 971
           | |||||||| |||| |||||||||| |||||||||||||||||||||| |||| ||||
Db     825 TAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATG 884

Qy     972 CAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAA 1031
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     885 CAAGAGCTCAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAA 944

Qy    1032 GGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTT 1091
           ||||||  || |||||  |||||||||||||||||||||||||||||||||||||||| |
Db     945 AGCTTCAGCTGCGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCT 1004

Qy    1092 GTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACC 1151
           |||||||||||||||  | ||||| |||||||||||| |||||| |||| |||| | |
Db    1005 GTGCTGTGTGCATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATC 1064

Qy    1152 ATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGT 1211
           |||||||||||||||||||| ||||| ||||| ||| |||||| ||||||||||||||
Db    1065 ATATTTTCCATAAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGT 1124
```

```
Qy   1212 GCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1271
          |||| |||||||| |||||||| ||||||||||||||||||||||||||||||||||||||
Db   1125 GCAAGTGTGACATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1184

Qy   1272 CTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATA 1331
          |||||||||| ||||| || |||||| |||||||| |||||  ||||||||||||| |
Db   1185 CTTTACAAGTTCCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACA 1244

Qy   1332 ATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTC 1391
          |||||| ||||| |||||||||||||||||||||||| ||| ||||||| || ||||
Db   1245 GTCGCAGTGAGACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTC 1304

Qy   1392 TGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATT 1451
          |||||||||| | ||||||| |||||||| ||||||||||||||||||||||||||||||
Db   1305 TGGAGGAACATGCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATT 1364

Qy   1452 CTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTC 1511
          ||||||| ||||||||| |||| ||||| |||||||||||||||||||||||| |||||||
Db   1365 CTGTGGCCGTGGATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTC 1424

Qy   1512 CTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTT-AAAATCTGTGTAAATAGAAAACT 1570
          || |||||||| | |||||||| || |||||||||| |||||||||| ||||||||||||
Db   1425 CTGATCAAGAGGCAGCTGTTCGGGAGATTAAATCTTAAAAATCTGTGTCAATAGAAAACT 1484

Qy   1571 TGAACCATTAG-TAATAACAGAACTGCCAATCAGGGCCTAGTTT-CTATTAATAAATTGG 1628
          |||||| |||| ||| ||||| |||||||||||||||||||| |||| ||| || |||
Db   1485 TGAACCGTTAGTTAACAACAGGACTGCCAATCAGGGCCTAGTTTACTATGAATGAACTGG 1544

Qy   1629 ATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATA 1688
          ||||     ||||| |||| |||||||||||||||| || | ||| ||||||| |||||
Db   1545 GTAAA---CGTAAAACAAGAATGATACTGAAAGTGCTGAGGTAACTTATATTATACTATA 1601

Qy   1689 GTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATA 1748
          ||||||   |  | || ||||||| ||   || | ||||  |  |||||||  ||||
Db   1602 GTTAAA--TGGCTTAACATATTTACCCCAGTACCGTTTT--CCACAAACTCACCATAACG 1657

Qy   1749 TTTTTCATAGGCAAGTTTCCTCT 1771
          |||||||||||||||||||||||
Db   1658 TTTTTCATAGGCAAGTTTCCTCT 1680
```

RESULT 5
AB041548
LOCUS AB041548 1936 bp mRNA linear ROD 30-JUN-2000
DEFINITION Mus musculus brain cDNA, clone MNCb-3816, similar to AF171875 g1-related zinc finger protein (Mus musculus).
ACCESSION AB041548
VERSION AB041548.1 GI:7670361
KEYWORDS fis (full insert sequence).
SOURCE Mus musculus (strain:C57BL) cDNA to mRNA, clone_lib:Sugano mouse brain mncb clone:MNCb-3816.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Osada,N., Kusuda,J., Tanuma,R., Ito,A., Hirata,M., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from mouse brain cDNA library made by oligo-capping method
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1936)
AUTHORS Hashimoto,K., Osada,N., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
COMMENT URL: http://www.nih.go.jp/yoken/genebank/
Lib Name: Sugano mouse brain mncb
Lab host: TOP10
Vector: pME18S-FL3
1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb.The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Library was constructed by Sugano et al.(University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTCTGCTCTAAAAGCTGCG]; 3' end primer [CGACCTGCAGCTCGAGCACA]).
A part of this sequence is reported in AU079081.
FEATURES Location/Qualifiers
source 1. .1936
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCb-3816"
/clone_lib="Sugano mouse brain mncb"
CDS 187. .1473
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA95033.1"
/db_xref="GI:7670362"
/translation="MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWT
AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNF
TVPTVWGSTVQVSWLALIQRGGGCTFADKVHLASERGASGAVIFNFPGTRNEVIPMSH
PGAGDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFII
TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDS
CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG
SVSLQVPVSNEASNTASPHEEDSRSETASSGYASVQGADEPPLEEHAQSANENLQLVN
HEANSVAVDVVPHVDNPTFEEDETPDQEAAVREIKS"
BASE COUNT 477 a 466 c 506 g 487 t
ORIGIN

Query Match 67.8%; Score 1203.6; DB 10; Length 1936;
Best Local Similarity 89.5%; Pred. No. 3.2e-265;
Matches 1363; Conservative 0; Mismatches 150; Indels 10; Gaps 6;

```
Qy    252 GCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCT 311
          | | |||||||||||||||||||| ||| |||||||| |||||||| ||||||||
Db    176 GAGTGCGCGCCATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAG 235

Qy    312 TTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCC 371
          | || |  | ||||| |||||||||| ||||| |||||||||| || ||||||||||
Db    236 CTGCCCGGCTACTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCC 295

Qy    372 GGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGG 431
          |||| || ||||| |||||||||||||||||||||||||||||||||||||||||||| |
Db    296 GGGGAGCCGAAGCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCG 355

Qy    432 GAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 491
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    356 GAGTGAACCGCACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 415

Qy    492 TGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTA 551
          ||||||| ||   | |||||||||||||||||||||||||||||||||| ||||||||||
Db    416 TGGAGCCCGTCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTA 475

Qy    552 ACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGT 611
          ||||||||| |||||||||||||||||||||||| ||||| |||||||| |||||||
Db    476 ACCCGCACACCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGT 535

Qy    612 TGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATG 671
          |||||||||||||||||| || |||||||||||||| |||||| |||||||||||  |
Db    536 TGGCCCTCATCCAACGCGGTGGAGGCTGCACCTTCGCGGACAAGGTCCATCTGGCTTCAG 595

Qy    672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
          ||||| |||| |||||||| |||||||||||||||| |||||||||||||||||||||||
Db    596 AGAGAGGGGCTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCC 655

Qy    732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
          |||||||||||||||||| |  |||||||||||||||||| |||||||||||||| |
Db    656 CCATGTCTCACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAA 715
```

```
Qy    792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
          |||||||||||||| ||||||||||||||||| ||||| |||||||||||| ||||||||||
Db    716 CAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGA 775

Qy    852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTA 911
          |||||||||||||||||||||||||| |||||||||| |||||||||||||||||||| |
Db    776 AAAAACATGGCCCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCA 835

Qy    912 TTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATG 971
          | |||||||| ||||| |||||||||| |||||||||||||||||||   |||| ||||
Db    836 TAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATG 895

Qy    972 CAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAA 1031
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    896 CAAGAGCTCAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAA 955

Qy   1032 GGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTT 1091
          |||||| || ||||  |||||||||||||||||||||||||||||||||||||||||| |
Db    956 AGCTTCAGCTGCGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCT 1015

Qy   1092 GTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACC 1151
          |||||||||||||||  | ||||| |||||||||||||| |||||| |||| ||||| |
Db   1016 GTGCTGTGTGCATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATC 1075

Qy   1152 ATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGT 1211
          |||||||||||||||||||||| |||| ||||| ||| |||||| ||||||||||||||
Db   1076 ATATTTTCCATAAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGT 1135

Qy   1212 GCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1271
          |||| |||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db   1136 GCAAGTGTGACATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1195

Qy   1272 CTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCCATGAAGAGGATA 1331
          |||||||||| ||||| || ||||||  |||||||| ||||||  |||||||||||||| |
Db   1196 CTTTACAAGTTCCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACA 1255

Qy   1332 ATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTC 1391
          |||||| |||||  |||||||||||||||||||||||||| ||| ||||||| || ||||
Db   1256 GTCGCAGTGAGACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTC 1315

Qy   1392 TGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATT 1451
          |||||||||| | ||||||| ||||||||| |||||||||||||||||||||||||||||
Db   1316 TGGAGGAACATGCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATT 1375

Qy   1452 CTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTC 1511
          ||||||| |||||||| |||| ||||| |||||||||||||||||||||| |||||||
Db   1376 CTGTGGCCGTGGATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTC 1435

Qy   1512 CTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTT-AAAATCTGTGTAAATAGAAAACT 1570
          || |||||||| | |||||||| || |||||||||| |||||||||| ||||||||||||
Db   1436 CTGATCAAGAGGCAGCTGTTCGGGAGATTAAATCTTAAAAATCTGTGTCAATAGAAAACT 1495

Qy   1571 TGAACCATTAG-TAATAACAGAACTGCCAATCAGGGCCTAGTTT-CTATTAATAAATTGG 1628
          |||||| |||| ||| ||||| |||||||||||||||||||| |||| ||| || |||
Db   1496 TGAACCGTTAGTTAACAACAGGACTGCCAATCAGGGCCTAGTTTACTATGAATGAACTGG 1555

Qy   1629 ATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATA 1688
          ||||     ||||| |||| |||||||||||||||| || | ||| ||||||| |||||
Db   1556 GTAAA---CGTAAAACAAGAATGATACTGAAAGTGCTGAGGTAACTTATATTATACTATA 1612

Qy   1689 GTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATA 1748
          ||||||   |  | || |||||||  ||   || | ||||  |  ||||||||  ||||
Db   1613 GTTAAA--TGGCTTAACATATTTACCCCAGTACCGTTTT--CCACAAACTCACCATAACG 1668

Qy   1749 TTTTTCATAGGCAAGTTTCCTCT 1771
          |||||||||||||||||||||||
Db   1669 TTTTTCATAGGCAAGTTTCCTCT 1691


RESULT 6
AF426411
LOCUS       AF426411               1287 bp    mRNA    linear   ROD 22-NOV-2001
DEFINITION  Mus musculus zinc ring finger-containing protein GRAIL mRNA,
            complete cds.
ACCESSION   AF426411
VERSION     AF426411.1  GI:17046405
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Ford,G.S., Bloom,D., Paragas,V., Anandasabapathy,N., Skrenta,H.,
            Ermann,J., Cron,R.Q., Lewis,D.B. and Fathman,G.
  TITLE     GRAIL: A novel gene expressed in anergic T cells that inhibits
            NFAT/AP-1 activation and IL-2 transcription
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1287)
  AUTHORS   Ford,G.S., Bloom,D., Paragas,V., Anandasabapathy,N., Skrenta,H.,
            Ermann,J., Cron,R.Q., Lewis,D.B. and Fathman,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) Medicine, Immunology and Rheumatology,
            Stanford University School of Medicine, 300 Pasteur Drive,
            Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1. .1287
                     /organism="Mus musculus"
                     /strain="DBA/2"
                     /db_xref="taxon:10090"
                     /clone="11.3.7"
                     /cell_type="anergic T cell"
     CDS             1. .1287
                     /codon_start=1
                     /product="zinc ring finger-containing protein GRAIL"
                     /protein_id="AAL34514.1"
                     /db_xref="GI:17046406"
                     /translation="MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWT
                     AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNF
                     TVPTVWGSTVQVSWLALIQRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSH
                     PGAGDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFII
                     TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDS
                     CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG
                     SVSLQVPVSNEASNTASPHEEDSRSETASSGYASVQGADEPPLEEHAQSANENLQLVN
                     HEANSVAVDVVPHVDNPTFEEDETPDQEAAVREIKS"
BASE COUNT      308 a    320 c    359 g    300 t
ORIGIN

  Query Match             62.4%;  Score 1106.8;  DB 10;  Length 1287;
  Best Local Similarity   91.2%;  Pred. No. 4.7e-243;
  Matches 1174;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    263 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
          ||||||||||||| |||  |||||||| |||||||| ||||||||   | || | | |
Db      1 ATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAGCTGCCCGGCTA 60

Qy    323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
          ||||| |||||||||||| ||||| |||||||||| || |||||||||||||| || |||
Db     61 CTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCCGGGGAGCCGAA 120

Qy    383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
          || ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    121 GCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCGGAGTGAACCGC 180

Qy    443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    181 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCCGTC 240

Qy    503 GCTGGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
          .  | |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    241 TCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTAACCCGCACACC 300

Qy    563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
          |||||||||||||||||||||||| ||||| |||||||| ||||||||||||||||||||
Db    301 AATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGTTGGCCCTCATC 360
```

```
Qy    623 CAACGCGGCGGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
          ||||||||  || ||||||||||||||| ||||||||||||||||||||  |||||| ||||
Db    361 CAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAGAGAGAGGGGCT 420

Qy    683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
          |||||||| ||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    421 TCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 480

Qy    743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
          ||||||||  |  |||||||||||||||||||| |||||||||||||| ||||||||||||
Db    481 CCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAACAAAAATTCTG 540

Qy    803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
          ||  |||||||||||||||| |||||| |||||||||||  ||||||||||||||||||||
Db    541 CAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGAAAAAACATGGC 600

Qy    863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
          ||||||||||||| ||||||||||| ||||||||||||||||||||  || |||||||||
Db    601 CCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTTCATAATTACGGCA 660

Qy    923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
          |||||  |||||||||||  ||||||||||||||||||||  |||| ||||||||||||||
Db    661 GCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATGCAAGAGCTCAA 720

Qy    983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
          |||||||||||||||| ||||||||||||||||||||||||||||||| |||||| ||
Db    721 AGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAAAGCTTCAGCTG 780

Qy   1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
          |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCTGTGCTGTGTGC 840

Qy   1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
          |||||  | |||||  |||||||||||||| ||||||  |||| |||||  ||||||||||||
Db    841 ATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATCATATTTTCCAT 900

Qy   1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
          ||||||||||  |||||  |||||  ||| ||||||  |||||||||||||||||  ||||||
Db    901 AAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGTGCAAGTGTGAC 960

Qy   1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
          ||  |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTT 1020

Qy   1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1342
          |||||  ||  ||||||  ||||||||  ||||||  |||||||||||||  |||||| |||
Db   1021 CCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACAGTCGCAGTGAG 1080

Qy   1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
          ||  ||||||||||||||||||||||||||  ||| |||||||  ||  ||||||||||||||
Db   1081 ACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTCTGGAGGAACAT 1140

Qy   1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
          | |||||||  ||||||||  ||||||||||||||||||||||||||||||||||||| |||
Db   1141 GCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCCGTG 1200

Qy   1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
          ||||||  ||||  |||||  |||||||||||||||||||||||  |||||||||  ||||||||
Db   1201 GATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTCCTGATCAAGAG 1260

Qy   1523 ACTGCTGTTCGAGAAATTAAATCTTAA 1549
            | ||||||||  ||  ||||||||||||
Db   1261 GCAGCTGTTCGGGAGATTAAATCTTAA 1287
```

```
RESULT 7
AK027169
LOCUS       AK027169               1871 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ23516 fis, clone LNG04848.
ACCESSION   AK027169
VERSION     AK027169.1  GI:10440232
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens human lung cDNA to mRNA, clone_lib:LNG clone:LNG04848.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
            Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
            Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M.,
            Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T.,
            Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1871)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .1871
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="LNG04848"
                     /tissue_type="human lung"
                     /clone_lib="LNG"
                     /note="cloning vector pME18SFL3"
     CDS             76. .546
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15682.1"
                     /db_xref="GI:10440233"
                     /translation="MEIVVLCALNCINQMIWYSILTCNHIFHKTCVDPWLLEHRTCPM
                     CKCDILKALGIEVDVEDGSASLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDE
                     PPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS"
BASE COUNT      670 a    277 c    321 g    603 t
ORIGIN
```

```
  Query Match             40.3%;  Score 714.4;  DB 9;  Length 1871;
  Best Local Similarity   97.9%;  Pred. No. 5e-153;
  Matches  755;  Conservative    0;  Mismatches   12;  Indels    4;  Gaps    3;

Qy   1005 AGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACA 1064
          ||||||||||||||||||  | |||||||||||||||| ||||||||||||||||||||
Db      1 AGGCAGATGCTAAAAAAGCGTTAGGAAGGCTTCAACTACTCACACTGAAACAAGGAGACA 60

Qy   1065 AGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATG 1124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATG 120

Qy   1125 ATTTGGTAC-GCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGG 1183
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTTGGTACAGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGG 180

Qy   1184 CTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATT 1243
          ||||||  ||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGTTAGAACACAGGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATT 240

Qy   1244 GAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCT 1303
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    241 GAGGTGGATGTTGAAGATGGATCAGCGTCTTTACAAGTCCCTGTATCCAATGAAATATCT 300

Qy   1304 AATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCT 1363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCT 360
```

```
Qy    1364 TCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGT 1423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGT 420

Qy    1424 CTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGAC 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     421 CTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGAC 480

Qy    1484 AACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAA 1543
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAA 540

Qy    1544 TCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAG 1603
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAG 600

Qy    1604 GGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTG 1663
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTG 660

Qy    1664 CTCAGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACT 1723
           ||||||||||||||||||||||||||||||    |  | ||||||||||||||||||||||
Db     661 CTCAGATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACT 718

Qy    1724 TTTTTCCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
           |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     719 TTTTTCCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 768
```

RESULT 8
AL391315
LOCUS AL391315 169812 bp DNA linear PRI 27-JUN-2001
DEFINITION Human DNA sequence from clone RP11-150F24 on chromosome X, complete sequence.
ACCESSION AL391315
VERSION AL391315.22 GI:14575271
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169812)
AUTHORS Brown,A.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 28, 2001 this sequence version replaced gi:14455902.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX
RP11-150F24 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-150F24 The true left end of clone RP11-321G1 is at 68861 in this sequence. The true right end of clone RP11-697G3 is at 161410 in this sequence.

FEATURES	Location/Qualifiers
source	1. .169812
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/clone="RP11-150F24"
	/clone_lib="RPCI-11.1"
misc_feature	12. .569
	/note="match: GSS: Em:AQ375615"
repeat_region	94. .165
	/note="2 copies 36 mer 90% conserved"
repeat_region	512. .729
	/note="MIR repeat: matches 3. .238 of consensus"
repeat_region	1338. .1638
	/note="AluY repeat: matches 1. .304 of consensus"
repeat_region	3599. .3642
	/note="L2 repeat: matches 2707. .2750 of consensus"
repeat_region	3722. .3899
	/note="MIR repeat: matches 12. .171 of consensus"
repeat_region	3915. .4028
	/note="MIR repeat: matches 131. .238 of consensus"
repeat_region	5121. .5197
	/note="L2 repeat: matches 2611. .2710 of consensus"
repeat_region	6734. .7029
	/note="AluSp repeat: matches 1. .297 of consensus"
repeat_region	7660. .7955
	/note="AluSq repeat: matches 1. .308 of consensus"
repeat_region	8891. .8940
	/note="L1PA9 repeat: matches 6114. .6163 of consensus"
repeat_region	10067. .10370
	/note="AluSx repeat: matches 1. .297 of consensus"
repeat_region	10411. .10695
	/note="AluSx repeat: matches 1. .288 of consensus"
repeat_region	10737. .10879
	/note="MER3 repeat: matches 5. .154 of consensus"
repeat_region	11078. .11380
	/note="AluY repeat: matches 1. .300 of consensus"
repeat_region	11571. .12560
	/note="L1MD2 repeat: matches 5328. .6331 of consensus"
repeat_region	13476. .13794
	/note="L1ME1 repeat: matches 5837. .6165 of consensus"
repeat_region	13906. .14041
	/note="L2 repeat: matches 2536. .2671 of consensus"
repeat_region	14106. .14411
	/note="AluY repeat: matches 2. .306 of consensus"
repeat_region	16413. .16704
	/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region	17566. .17697
	/note="FLAM_C repeat: matches 1. .132 of consensus"
repeat_region	17839. .18080
	/note="MIR repeat: matches 3. .260 of consensus"
repeat_region	18961. .19153
	/note="MIR repeat: matches 30. .262 of consensus"
repeat_region	19562. .19658
	/note="MER5A repeat: matches 12. .109 of consensus"
repeat_region	20326. .20413
	/note="L1MC4 repeat: matches 7861. .7948 of consensus"
repeat_region	20471. .20754
	/note="AluSq repeat: matches 1. .286 of consensus"
repeat_region	20755. .20782
	/note="14 copies 2 mer ta 89% conserved"
repeat_region	20999. .21184
	/note="L2 repeat: matches 2529. .2729 of consensus"
repeat_region	21194. .21952
	/note="L2 repeat: matches 1730. .2546 of consensus"
repeat_region	22817. .23035
	/note="MIR repeat: matches 15. .238 of consensus"
misc_feature	23641. .24552

```
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   24009. .24384
                     /note="L1M3e repeat: matches 310. .694 of consensus"
     misc_feature    complement(24313. .24743)
                     /note="match: GSS: Em:AQ923340"
     repeat_region   25462. .25711
                     /note="MIR repeat: matches 2. .256 of consensus"
     repeat_region   25838. .25939
                     /note="MIR repeat: matches 6. .106 of consensus"
     repeat_region   26420. .26566
                     /note="MIR repeat: matches 2. .151 of consensus"
     repeat_region   28132. .28302
                     /note="MIR repeat: matches 64. .238 of consensus"
     repeat_region   29706. .29835
                     /note="L2 repeat: matches 2606. .2746 of consensus"
     repeat_region   30126. .30458
                     /note="MER2 repeat: matches 2. .344 of consensus"
     repeat_region   30873. .31184
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   34418. .34503
                     /note="L1PA3 repeat: matches 6061. .6146 of consensus"
     repeat_region   34518. .34851
                     /note="MER2 repeat: matches 1. .344 of consensus"
     repeat_region   35308. .35457
                     /note="MIR repeat: matches 1. .162 of consensus"
     repeat_region   35828. .36039
                     /note="MIR repeat: matches 7. .226 of consensus"
     repeat_region   36274. .36349
                     /note="19 copies 4 mer ttcc 76% conserved"
     repeat_region   36353. .36642
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   37753. .37919
                     /note="FRAM repeat: matches -1. .165 of consensus"
     repeat_region   38928. .39134
                     /note="L2 repeat: matches 2467. .2703 of consensus"
     repeat_region   39127. .39319
                     /note="L2 repeat: matches 2031. .2228 of consensus"
     repeat_region   39546. .39726
                     /note="MER5A repeat: matches 1. .186 of consensus"
     repeat_region   39738. .40910
                     /note="Charliela repeat: matches 1. .1258 of consensus"
     repeat_region   40911. .41435
                     /note="L1PA5 repeat: matches 5616. .6143 of consensus"
     repeat_region   41436. .41674
                     /note="Charliela repeat: matches 1258. .207 of consensus"
     repeat_region   41675. .41942
                     /note="AluY repeat: matches 1. .283 of consensus"
     repeat_region   41943. .42098
                     /note="Charliela repeat: matches 207. .1455 of consensus"
     repeat_region   42634. .42699
                     /note="MSTB repeat: matches 356. .426 of consensus"
     repeat_region   43609. .43659
                     /note="L2 repeat: matches 2696. .2746 of consensus"
     repeat_region   45090. .45187
                     /note="L2 repeat: matches 2606. .2705 of consensus"
     repeat_region   45672. .46211
                     /note="L1MA4 repeat: matches 5700. .6295 of consensus"
     repeat_region   46808. .47107
                     /note="AluJb repeat: matches 1. .301 of consensus"
     repeat_region   47190. .47271
                     /note="HY1 repeat: matches 28. .111 of consensus"
     repeat_region   47355. .47455
                     /note="MER63 repeat: matches 5. .734 of consensus"
     repeat_region   47525. .47693
                     /note="MIR repeat: matches 3. .196 of consensus"
     misc_feature    47826. .47978
                     /note="match: GSS: Em:AQ136922"
     misc_feature    47985. .48294
                     /note="match: GSS: Em:AQ136922"
     repeat_region   48203. .50340
                     /note="L1M1 repeat: matches -1389. .1172 of consensus"
     repeat_region   50335. .51398
                     /note="L1PA2 repeat: matches 1679. .2716 of consensus"
     repeat_region   51399. .51743
                     /note="AluJb repeat: matches 1. .305 of consensus"
     repeat_region   51744. .52940
                     /note="L1PA2 repeat: matches 2716. .3982 of consensus"
     repeat_region   52917. .54922
                     /note="L1PA2 repeat: matches 4138. .6146 of consensus"
     repeat_region   54936. .57273
                     /note="L1MA3 repeat: matches 3951. .6304 of consensus"
     repeat_region   57453. .58059
                     /note="L1MB3 repeat: matches 5276. .5875 of consensus"
     repeat_region   58056. .58804
                     /note="L1PA10 repeat: matches 5104. .5833 of consensus"

  Query Match             38.0%;  Score 674.4;  DB 9;  Length 169812;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-143;
  Matches  675;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23909 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 23968

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23969 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 24028

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24029 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 24088

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24089 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 24148

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24149 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 24208

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24209 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 24268

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24269 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 24328

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24329 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 24388

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24389 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 24448

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24449 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 24508

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24509 AGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 24568

Qy     734 ATGTCTCACCCGGGTG 749
           ||||||||||||||||
Db   24569 ATGTCTCACCCGGGTG 24584


RESULT 9
AL591467
LOCUS       AL591467              169863 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome X clone RP11-697G3, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL591467
```

```
VERSION     AL591467.3  GI:14348539
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169863)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 12, 2001 this sequence version replaced gi:14268300.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA697G3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 157318 bases at least Q40
            Consensus quality: 157886 bases at least Q30
            Consensus quality: 158243 bases at least Q20
            Insert size: 169863; sum-of-contigs
            Insert size: 169949; 0.8% error; agarose-fp
            Quality coverage: 5.85x in Q20 bases; sum-of-contigs Quality
            coverage: 6.21x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .169863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-697G3"
                     /clone_lib="RPCI-11.3"
     misc_feature    1. .169863
                     /note="assembly_fragment:02868"
BASE COUNT     51891 a  31100 c  31417 g  55455 t
ORIGIN

  Query Match             38.0%;  Score 674.4;  DB 2;  Length 169863;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-143;
  Matches  675;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23909 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 23968

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23969 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 24028

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24029 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 24088

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24089 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 24148

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24149 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 24208

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24209 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 24268

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24269 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 24328

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24329 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 24388

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24389 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 24448

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24449 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 24508

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24509 AGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 24568

Qy     734 ATGTCTCACCCGGGTG 749
           ||||||||||||||||
Db   24569 ATGTCTCACCCGGGTG 24584


RESULT 10
AX071980
LOCUS       AX071980                404 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 2452 from Patent WO0102568.
ACCESSION   AX071980
VERSION     AX071980.1  GI:12582331
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitiz,D., Kita,D., Garcia,V. and Strache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 2452 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       134 a     81 c     90 g     99 t
ORIGIN

  Query Match             22.1%;  Score 392.2;  DB 6;  Length 404;
  Best Local Similarity   99.0%;  Pred. No. 2.8e-79;
  Matches  394;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    1194 ACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATG 1253
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7 AGAGGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATG 66

Qy    1254 TTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCT 1313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 TTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCT 126

Qy    1314 CCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGG 1373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 CCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGG 186

Qy    1374 GAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGG 1433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 GAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGG 246
```

```
Qy   1434 TAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCT 1493
          |||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db    247 TAAACCATGAAGCATATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCT 306

Qy   1494 TTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATC 1553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 TTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATC 366

Qy   1554 TGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGA 1591
          |||||||||||||||||||||||||||||||||||||
Db    367 TGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGA 404
```

RESULT 11
AX341568

```
LOCUS       AX341568                646 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 1815 from Patent WO0196388.
ACCESSION   AX341568
VERSION     AX341568.1  GI:18137550
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 1815 20-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      226 a    103 c    106 g    208 t      3 others
ORIGIN
```

```
  Query Match             21.0%;  Score 372.6;  DB 6;  Length 646;
  Best Local Similarity   98.0%;  Pred. No. 9.7e-75;
  Matches  398;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    2;

Qy   1369 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 1428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 60

Qy   1429 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCC 1488
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     61 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCC 120

Qy   1489 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 1548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 180

Qy   1549 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 1608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 240

Qy   1609 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 1668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 300

Qy   1669 ATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTT 1728
          |||||||||||||||||||||||||   |  | ||||||||||||||||||||||||||
Db    301 ATGACTAATATTATGTCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTT 358

Qy   1729 CCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          ||| |||||||||||||||||||||||||||||||||||||||||
Db    359 CCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 403
```

RESULT 12
AX396961/c

```
LOCUS       AX396961                435 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 1176 from Patent WO0212328.
ACCESSION   AX396961
VERSION     AX396961.1  GI:21067708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   King,G.E., Meagher,M.J., Xu,J. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0212328-A 1176 14-FEB-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      147 a     62 c     69 g    157 t
ORIGIN
```

```
  Query Match             17.5%;  Score 310;  DB 6;  Length 435;
  Best Local Similarity   97.4%;  Pred. No. 2e-60;
  Matches  336;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    2;

Qy   1430 CTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCA 1489
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    435 CTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCA 376

Qy   1490 ACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAA 1549
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    375 ACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGTTGTTCGAGAAATTAAATCTTAA 316

Qy   1550 AATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTA 1609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 AATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTA 256

Qy   1610 GTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGA 1669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGA 196

Qy   1670 TGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTC 1729
          ||||||||||||||||||||||||   |  | ||||||||||||||||||||||||||||
Db    195 TGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTC 138

Qy   1730 CACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          || ||||||||||||||||||||||||||||||||||||||||||
Db    137 CA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 94
```

RESULT 13
AC115936/c

```
LOCUS       AC115936              195263 bp    DNA     linear   HTG 20-JUN-2002
DEFINITION  Mus musculus clone RP24-545D3, WORKING DRAFT SEQUENCE, 13 ordered
            pieces.
ACCESSION   AC115936
VERSION     AC115936.3  GI:21490425
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 195263)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-545D3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195263)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
```

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 195263)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (20-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jun 20, 2002 this sequence version replaced gi:21431255.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L24918
Center clone name: 545_D_3
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190111 bases at least Q40
Consensus quality: 192623 bases at least Q30
Consensus quality: 193567 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 194063; sum-of-contigs
Quality coverage: 7.4 in Q20 bases; agarose-fp
Quality coverage: 7.4 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 1669: contig of 1669 bp in length
* 1670 1769: gap of 100 bp
* 1770 2684: contig of 915 bp in length
* 2685 2784: gap of 100 bp
* 2785 6907: contig of 4123 bp in length
* 6908 7007: gap of 100 bp
* 7008 9245: contig of 2238 bp in length
* 9246 9345: gap of 100 bp
* 9346 12194: contig of 2849 bp in length
* 12195 12294: gap of 100 bp
* 12295 20535: contig of 8241 bp in length
* 20536 20635: gap of 100 bp
* 20636 28882: contig of 8247 bp in length
* 28883 28982: gap of 100 bp
* 28983 43950: contig of 14968 bp in length
* 43951 44050: gap of 100 bp
* 44051 69397: contig of 25347 bp in length
* 69398 69497: gap of 100 bp
* 69498 91792: contig of 22295 bp in length
* 91793 91892: gap of 100 bp
* 91893 118569: contig of 26677 bp in length
* 118570 118669: gap of 100 bp
* 118670 150554: contig of 31885 bp in length
* 150555 150654: gap of 100 bp
* 150655 195263: contig of 44609 bp in length.

FEATURES Location/Qualifiers

source	1. .195263	
	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	/clone="RP24-545D3"	
	/clone_lib="RPCI-24 Male Mouse BAC"	
misc_feature	1. .1669	
	/note="assembly_fragment"	
misc_feature	1770. .2684	
	/note="assembly_fragment"	
misc_feature	2785. .6907	
	/note="assembly_fragment"	
misc_feature	7008. .9245	
	/note="assembly_fragment"	
misc_feature	9346. .12194	
	/note="assembly_fragment"	
misc_feature	12295. .20535	
	/note="assembly_fragment"	
misc_feature	20636. .28882	
	/note="assembly_fragment"	
misc_feature	28983. .43950	
	/note="assembly_fragment"	
misc_feature	44051. .69397	
	/note="assembly_fragment"	
misc_feature	69498. .91792	
	/note="assembly_fragment"	
misc_feature	91893. .118569	
	/note="assembly_fragment"	
misc_feature	118670. .150554	
	/note="assembly_fragment"	
misc_feature	150655. .195263	
	/note="assembly_fragment"	

BASE COUNT 58269 a 38652 c 38418 g 58724 t 1200 others
ORIGIN

Query Match 15.8%; Score 280.4; DB 2; Length 195263;
Best Local Similarity 60.8%; Pred. No. 4.6e-53;
Matches 546; Conservative 0; Mismatches 316; Indels 36; Gaps 4;

```
Qy    434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
          | |||||| | | |||  |||||| | |||   || || | |||     |||     ||
Db 179183 GGGAACCGCATGCTGTCAGAGCTGGGGGAGACCGGGGTCTTCGGAAGAAGCTCCATCTTG 179124
```

```
Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||   || || || ||  | || || || || ||          |  || || ||| |
Db 179123 AAGAGAGTAGCAGGAGTTGTTGTGCCACCAGAGGG---AAAAATTCAAAATGCTTGTGAT 179067

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||  | || | ||| |   |||   |                  |||   | ||| |
Db 179066 CCCAATACCACTTTTATCCTGCCACGG---------------AACAAAGAGCCCTGGATC 179022

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           || |||||  |||   | || || ||  ||||| || | || ||  |  |||| | ||||
Db 179021 GCACTCATTGAAC---GGGGAGGTTGTGCCTTCACACAGAAAATTAAGGTGGCATCTGAG 178965

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           | ||    ||||   |||||| ||| ||||| || ||  ||||  ||||  |||||
Db 178964 CACGGAGCCAGAGGAGTGATCATCTATAATTTCCCAGGTACTGGCAACCAGGTTTTCCCC 178905

Qy     734 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 793
           |||||||||| ||      | |||||| |||| | | ||||| || ||  | |||||
Db 178904 ATGTCTCACCAGGCGTTTGAAGACATCGTTGTAGTTATGATTGGTAACATCAAAGGTATG 178845

Qy     794 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 853
           | ||| |||| |  |||| |  || | || || ||| |  |  | || || || | |
Db 178844 GAGATTTTGCATTTAATTCGGAAGGGAGTTCATGTTACAGTCATGGTGGAGGTGGGAAGA 178785

Qy     854 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 913
           ||||||| |  |||| ||||||||||              || ||| |||||| | ||
Db 178784 AAACATGTCATTTGGCTGAATCACTA--------------TTTTGTCTCCTTTATGATC 178740

Qy     914 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 973
           | ||  | || ||  | |  ||||| | |||||||     ||| || ||  ||   |||
Db 178739 GTCACCACCGCCACGTTAGCATATTTCACCTTTTATCACATTCGGAGACTTTGGGTGGCA 178680

Qy     974 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAAGCTATTGGAAGG 1033
           |||  | | | |||||    || || |||||    |||    || |||||| ||||   |
Db 178679 AGAATTGAGAACAGGAGATGGAAGCGATTAACAAGAGAGCTCAAGAAAGCTTTTGGGCAG 178620

Qy    1034 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1093
           || ||| | ||   | | ||| |||| ||  | ||| ||   ||  |||  || || |||
Db 178619 CTGCAAGTTCGAGTATTAAAAGAAGGGGATGAAGAAGTTAATCCAAATGCTGACAGCTGT 178560

Qy    1094 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1153
           |   | ||| |||||   ||||| || |||||  | || || ||  | || ||||| |||
Db 178559 GTCATCTGCTTTGAAGCCTATAAGCCTAATGAAATAGTTCGGATTCTCACTTGCAAACAT 178500

Qy    1154 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGC 1213
           |||||| |||| ||  ||||||| ||| |  ||   |||  || ||||||||||||
Db 178499 TTTTTCCACAAGAATTGCATTGACCCCTGGATTCTAGCCCACGGCACATGCCCCATGTGC 178440

Qy    1214 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1273
           |||||||||| || |||||| |||| ||| || ||||  |||| |||||  |||  |||
Db 178439 AAATGTGACATTCTGAAAGCTCTGGGGATTCAGATGGACATTGAGGATGGGACAGACTCT 178380

Qy    1274 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATA 1331
           | ||||| |   | || |||||| |  |    |    ||  ||   ||||| || |
Db 178379 CTGCAAGTTCTGATGTCAAATGAATTGCCCGGAACCTTGTCACCCGTGGAAGAAGAGA 178322
```

```
RESULT 14
AC111461/c
LOCUS       AC111461               167021 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-213C21, *** SEQUENCING IN PROGRESS
            ***, 75 unordered pieces.
ACCESSION   AC111461
VERSION     AC111461.2  GI:21736011
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 167021)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167021)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 167021)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18701226.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GMJK
                Center clone name: CH230-213C21
            ---------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 89039 bases at least Q40
                Consensus quality: 93636 bases at least Q30
                Consensus quality: 97431 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
```

```
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1200: contig of 1200 bp in length
*     1201     1300: gap of unknown length
*     1301     2480: contig of 1180 bp in length
*     2481     2580: gap of unknown length
*     2581     4294: contig of 1714 bp in length
*     4295     4394: gap of unknown length
*     4395     5589: contig of 1195 bp in length
*     5590     5689: gap of unknown length
*     5690     7039: contig of 1350 bp in length
*     7040     7139: gap of unknown length
*     7140     8161: contig of 1022 bp in length
*     8162     8261: gap of unknown length
*     8262     9786: contig of 1525 bp in length
*     9787     9886: gap of unknown length
*     9887    11075: contig of 1189 bp in length
*    11076    11175: gap of unknown length
*    11176    12487: contig of 1312 bp in length
*    12488    12587: gap of unknown length
*    12588    13739: contig of 1152 bp in length
*    13740    13839: gap of unknown length
*    13840    15104: contig of 1265 bp in length
*    15105    15204: gap of unknown length
*    15205    16606: contig of 1402 bp in length
*    16607    16706: gap of unknown length
*    16707    18097: contig of 1391 bp in length
*    18098    18197: gap of unknown length
*    18198    19766: contig of 1569 bp in length
*    19767    19866: gap of unknown length
*    19867    21735: contig of 1869 bp in length
*    21736    21835: gap of unknown length
*    21836    22871: contig of 1036 bp in length
*    22872    22971: gap of unknown length
*    22972    24060: contig of 1089 bp in length
*    24061    24160: gap of unknown length
*    24161    25579: contig of 1419 bp in length
*    25580    25679: gap of unknown length
*    25680    27146: contig of 1467 bp in length
*    27147    27246: gap of unknown length
*    27247    28388: contig of 1142 bp in length
*    28389    28488: gap of unknown length
*    28489    30073: contig of 1585 bp in length
*    30074    30173: gap of unknown length
*    30174    31407: contig of 1234 bp in length
*    31408    31507: gap of unknown length
*    31508    33049: contig of 1542 bp in length
*    33050    33149: gap of unknown length
*    33150    34856: contig of 1707 bp in length
*    34857    34956: gap of unknown length
*    34957    36203: contig of 1247 bp in length
*    36204    36303: gap of unknown length
*    36304    38026: contig of 1723 bp in length
*    38027    38126: gap of unknown length
*    38127    39131: contig of 1005 bp in length
*    39132    39231: gap of unknown length
*    39232    40881: contig of 1650 bp in length
*    40882    40981: gap of unknown length
*    40982    42222: contig of 1241 bp in length
*    42223    42322: gap of unknown length
*    42323    43979: contig of 1657 bp in length
*    43980    44079: gap of unknown length
*    44080    45706: contig of 1627 bp in length
*    45707    45806: gap of unknown length
*    45807    47463: contig of 1657 bp in length
*    47464    47563: gap of unknown length
*    47564    49070: contig of 1507 bp in length
*    49071    49170: gap of unknown length
*    49171    50894: contig of 1724 bp in length
*    50895    50994: gap of unknown length
*    50995    52487: contig of 1493 bp in length
*    52488    52587: gap of unknown length
*    52588    54880: contig of 2293 bp in length
*    54881    54980: gap of unknown length
*    54981    57373: contig of 2393 bp in length
*    57374    57473: gap of unknown length
*    57474    58491: contig of 1018 bp in length
*    58492    58591: gap of unknown length
*    58592    59883: contig of 1292 bp in length
*    59884    59983: gap of unknown length
*    59984    61364: contig of 1381 bp in length
*    61365    61464: gap of unknown length
*    61465    62800: contig of 1336 bp in length
*    62801    62900: gap of unknown length
*    62901    64401: contig of 1501 bp in length
*    64402    64501: gap of unknown length
*    64502    66826: contig of 2325 bp in length
*    66827    66926: gap of unknown length
*    66927    67935: contig of 1009 bp in length
*    67936    68035: gap of unknown length
*    68036    70512: contig of 2477 bp in length
*    70513    70612: gap of unknown length
*    70613    72597: contig of 1985 bp in length
*    72598    72697: gap of unknown length
*    72698    74479: contig of 1782 bp in length
*    74480    74579: gap of unknown length
*    74580    76830: contig of 2251 bp in length
*    76831    76930: gap of unknown length
*    76931    79798: contig of 2868 bp in length
*    79799    79898: gap of unknown length
*    79899    82654: contig of 2756 bp in length
*    82655    82754: gap of unknown length
*    82755    84685: contig of 1931 bp in length
*    84686    84785: gap of unknown length
*    84786    87419: contig of 2634 bp in length
*    87420    87519: gap of unknown length
*    87520    89991: contig of 2472 bp in length
*    89992    90091: gap of unknown length
*    90092    91827: contig of 1736 bp in length

  Query Match           15.7%;  Score 278.4;  DB 2;  Length 167021;
  Best Local Similarity 60.7%;  Pred. No. 1.3e-52;
  Matches  544;  Conservative    0;  Mismatches  316;  Indels   36;  Gaps    4;

Qy    434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
          | |||||| | | |||  |||||| |||||   || || | |||      |||      ||
Db  83860 GGGAACCGCATGCTGTCAGAGCTGGGCGAGACTGGGGTCTTCGGAAGAAGCTCCATCTTG 83801

Qy    494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
          ||   || || || ||  | || || || || || ||           |  || || ||| ||
Db  83800 AAGAGAGTAGCAGGAGTAGTTGTGCCACCAGAGGGAAA---AATTCAAAATGCTTGTGAC 83744

Qy    554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
          ||  | || | |||||   ||||  |                  |||   | ||| |
Db  83743 CCCAATACCAGTTTCATCCTGCCGCGGAA---------------CAAAGAGCCGTGGATT 83699

Qy    614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
          || |||||  ||    | || || ||  ||||| || | || ||| |  |||| | ||||
Db  83698 GCACTCATTGAA---AGGGGAGGTTGTGCCTTCACACAGAAAATCAAAGTGGCATCTGAG 83642

Qy    674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
          |  ||||    ||||  |||||| |||||| || |||||  |||| |||| |||||
Db  83641 AATGGGGCCAGGGGAGTGATCATCTATAACTTTCCGGGGACTGGCAACCAGGTTTTCCCC 83582

Qy    734 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 793
          |||||||||| ||      | |||||| |||| | | ||||| || ||  | ||||| |
Db  83581 ATGTCTCACCAGGCGTTTGAAGACATCGTTGTAGTTATGATTGGTAACGTCAAAGGTATG 83522

Qy    794 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 853
          | ||| |||| |  |||  |  ||  | || || ||  |  |  | || || || | |
Db  83521 GAGATTTTGCATTTAATTAGGAAGGGAGTCCATGTTACGGTCATGGTTGAGGTGGGAAGA 83462
```

```
Qy     854 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 913
           ||||| | |  |||| ||||||| ||              || ||| |||||| | ||
Db   83461 AAACACGTCATTTGGCTGAATCATTA--------------TTTTGTCTCCTTTATGATC 83417

Qy     914 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 973
             | ||  | || ||  | |  |||||  |  |||||   |  ||| || ||  ||   |||
Db   83416 GTCACAACCGCCACATTAGCATATTTCACTTTTTACCATATTCGGAGACTCTGGGTGGCA 83357

Qy     974 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 1033
           |||    |  |||||   ||  | |||||    |||    || |||||  ||||   |
Db   83356 AGAATCGAGGACAGGAGATGGAAACGATTAACAAGAGAGCTCAAGAAAGCATTTGGGCAG 83297

Qy    1034 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1093
           || ||| | || | | | ||| | || ||  ||||| |  | ||  ||| ||| || ||
Db   83296 CTGCAAGTTCGAATATTAAAAGAGGGGGATGAGGAAGTAAGTCCAAATGCAGACAGCTGC 83237

Qy    1094 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1153
           |   | ||| |||||   ||||| || |||||  | || || ||  | ||||||| |||
Db   83236 GTCATCTGCTTTGAAGCCTATAAGCCTAATGAAATAGTTCGTATTCTCACGTGCAAACAT 83177

Qy    1154 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGC 1213
           |||||||  ||||  ||  |||||||  |||  |  ||  ||   || ||||||||||||
Db   83176 TTTTTCCACAAGAATTGCATTGACCCCTGGATTCTAGCCCATGGCACATGCCCCATGTGC 83117

Qy    1214 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1273
           ||||| ||||| || |||||| |||| ||| || |||||  |||| ||||||||||  |||
Db   83116 AAATGCGACATTCTGAAAGCTCTGGGGATTCAGATGGACATTGAGGATGGATCAGACTCT 83057

Qy    1274 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGA 1329
           | ||||| |   | || ||||||| |  ||   |      ||  |    ||||| ||
Db   83056 CTGCAAGTTCTGATGTCAAATGAATTGCCTGGAACCTTTTCAGCTATGGAAGAAGA 83001
```

RESULT 15
AB070023
LOCUS AB070023 1391 bp mRNA linear PRI 16-AUG-2001
DEFINITION Macaca fascicularis testis cDNA clone:QtsA-11567, full insert sequence.
ACCESSION AB070023
VERSION AB070023.1 GI:15207886
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Macaca fascicularis adult male testis cDNA to mRNA, clone_lib:macaque testis cDNA library QtsA clone:QtsA-11567.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE 1
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J., Tanuma,R., Hirai,M., Terao,K. and Sugano,S.
TITLE Isolation of novel full-length cDNA clones from macaque testis cDNA libraries
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1391)
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
COMMENT Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb.The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al.,University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTCTGCTCTAAAAGCTGCG]; 3' end primer [CGACCTGCAGCTCGAGCACA]).
FEATURES Location/Qualifiers
source 1. .1391
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QtsA-11567"
/sex="male"
/tissue_type="testis"
/clone_lib="macaque testis cDNA library QtsA"
/dev_stage="adult"
CDS 198. .1328
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB62968.1"
/db_xref="GI:15207887"
/translation="MHLLKVGTWRNNTAFSWLIMFGVLWLVSQNCCRASVVWTAYMNI SFHVGNHVLSELGETGVFGRSSTLKRVAGVIVPPEGKIQNACNPNTIFSRSKYSETWL ALIERGGCTFTQKIKVAAEKGASGVIIYNFPGTGNQVFPMFHQAFEDVVVVMIGNLKG TEIFHLIKKGVLITAMVEVGRKHIIWMNHYLVSFVIVTTATLAYFIFYHIHRLCLARI QNRRWQRLTTDLQNAFGQQLQLRVVKEGDEEINPNGDSCVICFEHYKPNDIVRILTCKH FFHKNCIDPWILSHGTCPICKCDILKVLGIQVDVENGTEPLQVLMSSELCETLSPSEE ETNNEVSPAGTSDKVIHVEENPTSQNNDSQPHSVVEDVHPSP"
BASE COUNT 419 a 244 c 312 g 416 t
ORIGIN

Query Match 15.2%; Score 268.8; DB 9; Length 1391;
Best Local Similarity 62.5%; Pred. No. 7.3e-51;
Matches 445; Conservative 0; Mismatches 252; Indels 15; Gaps 1;

```
Qy     630 GCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCGTCTGGAG 689
           | || || || |||||| || | || ||  |  ||||   |||||   | ||   |||||
Db     517 GGGGAGGTTGTACCTTCACACAGAAAATTAAAGTGGCAGCTGAGAAGGGAGCCAGTGGAG 576

Qy     690 CCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTG 749
           |||||| |||||| || || ||  ||||  ||||  |||||||||| ||| | ||
Db     577 TGATCATCTATAACTTTCCAGGAACTGGCAACCAGGTGTTCCCCATGTTTCATCAGGCAT 636

Qy     750 CAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTGCAATCTA 809
             | |||  | ||||   | ||||| || ||  | ||||||||  ||||| | || |  |
Db     637 TTGAAGATGTCGTTGTGGTTATGATTGGTAACTTAAAAGGCACGGAAATTTTCCATTTAA 696

Qy     810 TTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGCCCTTGGG 869
           || | |  ||  | |  | |||   |  | || || |||| ||| ||  |  |||
Db     697 TTAAGAAGGGAGTTCTCATTACAGCCATGGTTGAGGTGGGGAGAAAGCACATCATCTGGA 756

Qy     870 TGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCGGCAACTG 929
           ||||||||||                ||  || || ||| | ||| | || | || ||
Db     757 TGAATCACTA--------------TTTGGTCTCTTTTGTGATTGTCACAACTGCTACCT 801

Qy     930 TGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAAAGCAGGA 989
           | |  ||||| |||||||||      ||  || ||  |    || ||   ||| | | |||
Db     802 TAGCATATTTCATCTTTTATCACATTCATAGACTTTGTTTAGCGAGGATTCAGAACCGGA 861

Qy     990 AGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTACGCACAC 1049
             | || |||||    |||||  | | || ||  |||||    || |||| ||
Db     862 GATGGCAGCGATTAACAACAGATCTTCAGAATGCATTTGGACAACTCCAACTTCGAGTGG 921

Qy    1050 TGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAAT 1109
           | ||| | || ||  ||||||   ||  ||||||||| || |   | ||| |||||
Db     922 TAAAAGAGGGGGATGAGGAAATAAATCCAAATGGAGATAGCTGCGTAATTTGCTTTGAAC 981

Qy    1110 TGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACAT 1169
           |||||| || |||||  | || || ||  | || || || ||| ||||||| ||||  |
Db     982 ACTATAAGCCTAATGACATAGTTCGTATTCTGACTTGTAAACATTTTTTCCACAAGAATT 1041

Qy    1170 GTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCA 1229
           | ||||||| ||| | |||   ||   ||| |||||||| |||||||||| || || |
Db     1042 GCATTGACCCCTGGATTTTATCCCATGGGACATGCCCCATTTGCAAATGTGATATTCTAA 1101
```

```
Qy    1230 AAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTAT 1289
           ||| |||||| ||| | ||||||||||||| ||||| |||    |||| ||||| |   | |
Db    1102 AAGTTTTGGGGATTCAAGTGGATGTTGAAAATGGAACAGAACCTTTGCAAGTTCTAATGT 1161

Qy    1290 CCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGA 1341
           | | |||| | | | | |      ||  |   ||||||||| |     | |||
Db    1162 CAAGTGAACTGTGTGAAACCTTATCACCTAGTGAAGAGGAGACAAATAACGA 1213
```

Search completed: April 26, 2003, 02:04:58
Job time : 3695 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 00:41:20 ; Search time 284 Seconds
(without alignments)
14067.060 Million cell updates/sec

Title: US-09-854-300-7
Perfect score: 1774
Sequence: 1 agctggagctccaccgcggt.....ataggcaagtttcctctcag 1774

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1773	99.9	1774	24	AAI72095	Human GRAIL cDNA.
	2	1662.8	93.7	2773	21	AAA15995	Human protein clon
	3	1662.8	93.7	2796	24	AAI72321	ISIGP-4 cDNA. Hom
	4	1431.4	80.7	2876	22	AAS42511	Human cDNA encodin
	5	1278.2	72.1	1284	21	AAA15985	Human protein clon
	6	1178.8	66.4	2306	22	AAS25884	Human cDNA encodin
	7	1167.2	65.8	1249	24	ABL90796	Human polynucleoti
	8	1166	65.7	2145	24	AAI72094	Mouse GRAIL cDNA.
	9	1157.6	65.3	1250	22	AAS26340	Human cDNA encodin
	10	1106.8	62.4	1287	24	ABK12983	Mouse goliath cDNA
	11	1052.4	59.3	1287	24	ABK12992	Synthetic goliath
	12	1042.8	58.8	1287	24	ABK12990	Synthetic goliath
	13	1042.8	58.8	1287	24	ABK12991	Synthetic goliath
	14	1019.4	57.5	1135	22	AAI59292	Human polynucleoti
	15	999.6	56.3	1287	24	ABK12989	Synthetic goliath
	16	977.2	55.1	1287	24	ABK12987	Synthetic goliath
	17	972.4	54.8	1287	24	ABK12988	Synthetic goliath
	18	655.2	36.9	737	22	AAI61078	Human polynucleoti
c	19	497.2	28.0	502	24	ABN19991	Human ORFX polynuc
	20	392.2	22.1	404	22	AAF66696	Novel human polynu
	21	372.6	21.0	646	24	ABL38226	Human colon tumour
c	22	310	17.5	435	24	ABK45625	cDNA encoding colo
	23	254.8	14.4	1355	21	AAA87765	Human secreted pro
	24	254.8	14.4	1355	22	AAF64047	cDNA encoding huma
	25	243.8	13.7	918	23	AAS70124	DNA encoding novel
	26	210.2	11.8	2826	24	ABL90797	Human polynucleoti
	27	209	11.8	1200	22	AAF94484	Human hydrophobic
	28	209	11.8	1527	22	AAF94494	Human hydrophobic
	29	209	11.8	2005	22	AAK94499	Human full-length
	30	209	11.8	2063	19	AAV43601	Human secreted pro
	31	209	11.8	2452	22	AAK94373	Human full-length
	32	194.8	11.0	1869	22	ABA09101	Human secreted pro
	33	194.8	11.0	1869	22	AAK51507	Human polynucleoti
	34	194.8	11.0	1869	22	AAK52491	Human polynucleoti
	35	191.8	10.8	1708	23	AAS76863	DNA encoding novel
	36	180.8	10.2	3615	24	ABK36205	cDNA sequence #596
	37	178.8	10.1	631	21	AAC76002	Human ORFX ORF1557
	38	149.2	8.4	1322	19	AAV43616	Human secreted pro
	39	148.8	8.4	1260	24	ABK12986	Mouse g1 cDNA sequ
	40	147.6	8.3	1253	19	AAV40387	Human zinc binding
	41	147.6	8.3	1257	22	AAF94418	Human hydrophobic
	42	147.6	8.3	1433	22	AAF94428	Human hydrophobic
	43	147.6	8.3	1489	24	ABL55072	Human cDNA sequenc
	44	147.6	8.3	1515	24	ABL90798	Human polynucleoti
	45	147.6	8.3	1546	24	ABL55094	Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAI72095
ID AAI72095 standard; cDNA; 1774 BP.
XX
AC AAI72095;
XX
DT 25-MAR-2002 (first entry)
XX
DE Human GRAIL cDNA.
XX
KW Murine; human; GRAIL; anergy; attenuation; tyrosine; phosphorylation;
KW antigenic stimulation; interleukin-2; gene therapy; polymorphism; IL-2;
KW autoimmune disease; tumour cell; cancer; transplant rejection; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 263..1547
FT /*tag= a
FT /product= "GRAIL"
XX
PN WO200185943-A1.
XX
PD 15-NOV-2001.
XX
PF 11-MAY-2001; 2001WO-US15385.
XX
PR 11-MAY-2000; 2000US-203513P.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Ford GS, Bloom D, Fathman CG;

XX
DR WPI; 2002-055597/07.
DR P-PSDB; AAB47800.
XX
PT Novel nucleic acid sequences of anergy associated genes, including
PT GRAIL gene useful in the evaluation of pathophysiology or immunotherapy
PT of cancer, autoimmune disease and transplant rejection -
XX
PS Claim 2; Page 45-47; 50pp; English.
XX
CC This sequence encodes human GRAIL protein. The GRAIL gene is an
CC anergy associated gene which is upregulated during the early stages of
CC induction of anergy. GRAIL has been shown to attenuate IL-2
CC transcription in T-cells during response to antigenic stimulation.
CC GRAIL DNA is useful for decreasing the responsiveness of a T cell
CC population, especially synthesis of interleukin-2 (IL-2) in response to
CC antigenic stimulation, by up-regulating GRAIL activity in the T cell
CC population. GRAIL DNA is useful in producing compositions that modulate
CC induction or maintenance of anergy, for gene therapy, mapping
CC functional regions of the encoded protein, to analyse a patient sample
CC for the presence of polymorphisms or alterations in expression of
CC sequences associated with T cell anergy, disease states, genetic
CC predisposition to a disease state, and in studying associated
CC physiological pathways. Modulation of the gene activity in vivo is
CC useful for prophylactic and therapeutic purposes, such as treating
CC autoimmune disease and to enhance immune response to tumour cells, and
CC identification of anergic T cells. GRAIL DNA is useful in the
CC evaluation of the pathophysiology or immunotherapy of cancer,
CC autoimmune disease, and transplant rejection. Genetic sequences
CC involved in anergy induction are useful as markers in the evaluation of
CC specific immunotherapies. Functional characterization of genes involved
CC in anergy induction allows the elucidation of the mechanisms of T cell
CC anergy, including the transcriptional blockade of IL-2, which may be
CC manipulated to regulate T cell responses in human disease. Modulation
CC of expression of GRAIL gene is useful in manipulating the anergic
CC state. The genetic sequences find use alone or in combinations in
CC determining the expression profile of cells relating to anergy, for
CC example in screening of candidate biologically active compounds for
CC modulation of T cell anergy. Expression of GRAIL has been found to
CC block tyrosine phosphorylation of a protein present in T cells during
CC activation.
XX
SQ Sequence 1774 BP; 454 A; 432 C; 464 G; 423 T; 1 other;

Query Match 99.9%; Score 1773; DB 24; Length 1774;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1774; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60

Qy      61 AATTCGGCACGAGCCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AATTCGGCACGAGCCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCT 120

Qy     121 CCTGGCTCCGACGTAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCTGGCTCCGACGTAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGC 180

Qy     181 GCACCTGCTCAAGACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCACCTGCTCAAGACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAG 240

Qy     241 GGAACTGCGGAGCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGAACTGCGGAGCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGG 300

Qy     301 TGGCTGCGGCTTTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGGCTGCGGCTTTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGC 360

Qy     361 ACCCGGTTCCCGGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACCCGGTTCCCGGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGT 420

Qy     421 TCCGCACACGGGAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCCGCACACGGGAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCA 480

Qy     481 GGACTCGCCGCTGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGACTCGCCGCTGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCT 540

Qy     541 TAACGCCTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TAACGCCTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCA 600

Qy     601 AGTCTCTTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGTCTCTTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCA 660

Qy     661 TCTGGCTTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TCTGGCTTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAA 720

Qy     721 TGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAA 780

Qy     781 TCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCAT 840

Qy     841 AGAAGTAGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGAAGTAGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGT 900

Qy     901 GTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAG 960

Qy     961 GCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAA 1020

Qy    1021 AGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGA 1080

Qy    1081 TGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTT 1140

Qy    1141 AACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGAC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGAC 1200

Qy    1201 TTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGA 1260

Qy    1261 TGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCA 1320

Qy    1321 TGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGA 1380

Qy    1381 TGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCA 1440

Qy    1441 TGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1441 TGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGA 1500

Qy   1501 AGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAA 1560

Qy   1561 ATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAA 1620

Qy   1621 TAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATT 1680

Qy   1681 ATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCA 1740

Qy   1741 TTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          |||||||||||||||||||||||||||||||||
Db   1741 TTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
```

RESULT 2
AAA15995
ID AAA15995 standard; cDNA; 2773 BP.
XX
AC AAA15995;
XX
DT 12-JUN-2000 (first entry)
XX
DE Human protein clone HP10574 full length coding sequence.
XX
KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW coagulation disorder; myocardial infarction; inflammatory condition;
KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW nephritis; therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200005367-A2.
XX
PD 03-FEB-2000.
XX
PF 22-JUL-1999; 99WO-JP03929.
XX
PR 24-JUL-1998; 98JP-0208820.
PR 07-AUG-1998; 98JP-0224105.
PR 25-AUG-1998; 98JP-0238116.
PR 09-SEP-1998; 98JP-0254736.
PR 29-SEP-1998; 98JP-0275505.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2000-182694/16.
DR P-PSDB; AAY94897.
XX
PT Novel human proteins having hydrophobic domains useful for treating
PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS Claim 4; Page 348-351; 351pp; English.
XX

CC This sequence encodes a human protein of the invention, which has
CC hydrophobic domains. The DNA sequences can be used as a probe or as a
CC genetic marker. The protein can also be used as a marker, and to identify
CC potential genetic disorders. The DNA and protein can also be used as
CC nutritional sources or supplements. The protein exhibits cytokine, cell
CC proliferation, cell differentiation activities and induces production of
CC other cytokines in certain cell populations. The protein also exhibits
CC immune stimulating or immune suppressing activity. It can be used in the
CC treatment of various immune deficiencies and disorders, and to treat
CC infectious diseases caused by viral, bacterial, fungal or other
CC infections. The protein is also used for treating autoimmune disorders
CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC arthritis. It is also useful in the treatment of allergic reactions and
CC conditions such as asthma, and in immune suppression after organ
CC transplantation. The protein is useful in regulation of haematopoiesis
CC and consequently in the treatment of myeloid or lymphoid cell
CC deficiencies. It is also used in compositions for tissue growth or
CC regeneration. The protein is also used in the treatment of osteoporosis
CC or osteoarthritis and in the treatment of periodontal disease and other
CC tooth repair processes. The protein is used in the treatment of nervous
CC system disorders such as Alzheimer's disease, Parkinson's disease, and
CC Huntington's disease. They are useful for protection or regeneration and
CC treatment of lung or liver fibrosis, reperfusion injury in various
CC tissues, and conditions resulting from systemic cytokine damage. They are
CC also used for promoting or inhibiting tissue differentiation. They are
CC also used as contraceptives since they exhibit activin or inhibin related
CC activities and as a fertility inducing therapeutic. They are used for
CC treating various coagulation disorders and in treatment and prevention of
CC conditions resulting from coagulation activities e.g. myocardial
CC infarction or stroke. They also acts as receptors, receptor ligands or
CC inhibitors or agonists of receptor/ligand interactions. They are used to
CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC reperfusion injury, arthritis, and nephritis. They can be used to
CC prevent tumours.
XX
SQ Sequence 2773 BP; 803 A; 545 C; 617 G; 808 T; 0 other;

Query Match 93.7%; Score 1662.8; DB 21; Length 2773;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1690; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

```
Qy     74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 81

Qy    134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 141

Qy    194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 201

Qy    254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 261

Qy    314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 321

Qy    374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 381

Qy    434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 441

Qy    494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 501
```

```
Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     502 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 561

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     562 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 621

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     622 AGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 681

Qy     734 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     682 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 741

Qy     794 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     742 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 801

Qy     854 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     802 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 861

Qy     914 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     862 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 921

Qy     974 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     922 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 981

Qy    1034 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     982 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1041

Qy    1094 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1042 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1101

Qy    1154 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGC 1213
           ||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db    1102 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTGC 1161

Qy    1214 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1162 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1221

Qy    1274 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1222 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1281

Qy    1334 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1282 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1341

Qy    1394 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1342 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1401

Qy    1454 GTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1513
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    1402 GTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1461

Qy    1514 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1462 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1521

Qy    1574 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1522 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1581

Qy    1634 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1582 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1641

Qy    1694 AATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATATTTTT 1753
           |   |  | ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1642 A--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATAATATTTTT 1698

Qy    1754 CATAGGCAAGTTTCCTCTCAG 1774
           ||||||||||||||||||||
Db    1699 CATAGGCAAGTTTCCTCTCAG 1719
```

RESULT 3
AAI72321

ID AAI72321 standard; cDNA; 2796 BP.
XX
AC AAI72321;
XX
DT 15-APR-2002 (first entry)
XX
DE ISIGP-4 cDNA.
XX
KW Human; intracellular signalling protein; ISIGP; gene; cell proliferation;
KW autoimmune; inflammation; gastrointestinal disorder;
KW reproductive disorder; developmental disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 217..1503
FT /*tag= a
FT /product= "ISIGP-4"
XX
PN WO200194391-A2.
XX
PD 13-DEC-2001.
XX
PF 07-JUN-2001; 2001WO-US18595.
XX
PR 08-JUN-2000; 2000US-210582P.
PR 16-JUN-2000; 2000US-212443P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, He A, Nguyen DB, Yao MG, Bandman O, Burford N, Tang YT;
PI Xu Y, Hafalia A, Azimzai Y, Walia NK;
XX
DR WPI; 2002-154564/20.
DR P-PSDB; AAB47874.
XX
PT New human intracellular signaling protein and polynucleotides useful
PT for diagnosing, treating or preventing cell proliferative,
PT autoimmune/inflammatory, gastrointestinal, reproductive and
PT developmental disorders -
XX
PS Claim 5; Page 104; 106pp; English.
XX
CC The sequences given in AAI72318-22 encode novel human intracellular
CC signalling proteins (ISIGP). The polynucleotides and ISIGP proteins
CC may be used for the diagnosis, treatment or prevention of cell
CC proliferative, autoimmune/inflammatory, gastrointestinal, reproductive
CC and developmental disorders. The protein encoded by this sequence has
CC homology to mouse gl-related zinc finger protein.
XX
SQ Sequence 2796 BP; 821 A; 548 C; 620 G; 807 T; 0 other;

Query Match 93.7%; Score 1662.8; DB 24; Length 2796;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1690; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

```
Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 87
```

```
Qy    134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     88 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 147

Qy    194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    148 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 207

Qy    254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    208 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 267

Qy    314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 327

Qy    374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    328 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 387

Qy    434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    388 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 447

Qy    494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    448 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 507

Qy    554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    508 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 567

Qy    614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    568 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 627

Qy    674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    628 AGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 687

Qy    734 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    688 ATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACA 747

Qy    794 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    748 AAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAA 807

Qy    854 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    808 AAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATT 867

Qy    914 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 973
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    868 ATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCA 927

Qy    974 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 1033
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    928 AGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGG 987

Qy   1034 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1093
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    988 CTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGT 1047

Qy   1094 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1153
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1048 GCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCAT 1107

Qy   1154 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGC 1213
          |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db   1108 ATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTGC 1167

Qy   1214 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1273
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1168 AAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCT 1227

Qy   1274 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1333
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1228 TTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAAT 1287

Qy   1334 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1288 CGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTG 1347

Qy   1394 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1348 GAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCT 1407

Qy   1454 GTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1513
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1408 GTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCCT 1467

Qy   1514 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1468 AATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGA 1527

Qy   1574 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1633
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1528 ACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAA 1587

Qy   1634 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1588 TTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAA 1647

Qy   1694 AATGGCTTAAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATATTTTT 1753
          |   |  | ||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1648 A--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATAATATTTTT 1704

Qy   1754 CATAGGCAAGTTTCCTCTCAG 1774
          ||||||||||||||||||||
Db   1705 CATAGGCAAGTTTCCTCTCAG 1725
```

RESULT 4
AAS42511
ID AAS42511 standard; cDNA; 2876 BP.
XX
AC AAS42511;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human cDNA encoding an mddt protein, clone LI:757439.1:2000MAY01.
XX
KW Human; molecules for disease detection and treatment; mddt; ss;
KW Antiarteriosclerotic; hepatotropic; antipsoriatic; cytostatic;
KW immunosuppressive; antidiabetic; antiasthmatic; neuroprotective;
KW osteopathic; antiarthritic; cell proliferative disorder;
KW arteriosclerosis; cirrhosis; psoriasis; cancer; adenocarcinoma;
KW leukaemia; breast cancer; autoimmune disorder; AIDS;
KW acquired immunodeficiency syndrome; Addison's disease;
KW diabetes mellitus; asthma; multiple sclerosis; osteoarthritis.
XX
OS Homo sapiens.
XX
PN WO200162922-A2.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US05896.
XX
PR 24-FEB-2000; 2000US-0185213.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.

PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR WPI; 2001-570631/64.
DR P-PSDB; AAU25459.
XX
PT New disease detection and treatment molecule polynucleotides and
PT polypeptides, useful for diagnosis and treatment of arteriosclerosis,
PT cirrhosis, psoriasis, cancer, autoimmune disorders, diabetes mellitus,
PT asthma and multiple sclerosis -
XX
PS Claim 1; Page 134-135; 183pp; English.
XX
CC The invention relates to novel human molecules for disease
CC detection and treatment (mddt proteins) and the polynucleotides encoding
CC them. The MDDT polynucleotides and polypeptides are useful for diagnostic
CC and therapeutic purposes e.g. to diagnose and treat cell proliferative
CC disorders (e.g. arteriosclerosis, cirrhosis and psoriasis) cancers (e.g.
CC adenocarcinoma, leukaemia and breast cancer) autoimmune disorders
CC (e.g. acquired immunodeficiency syndrome (AIDS) and Addison's disease)
CC diabetes mellitus, asthma, multiple sclerosis, osteoarthritis, and many
CC more diseases given in the specification. The present sequence
CC encodes an mddt protein of the invention.
XX
SQ Sequence 2876 BP; 840 A; 566 C; 642 G; 825 T; 3 other;

Query Match 80.7%; Score 1431.4; DB 22; Length 2876;
Best Local Similarity 96.5%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 28; Indels 33; Gaps 19;

```
Qy     74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGC-TCCGAC 132
Db     33 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTTCCGAC 92

Qy    133 GTAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAA 192
Db     93 GTAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAA 152

Qy    193 GACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAG 252
Db    153 GACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAG 212

Qy    253 CGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTT 312
Db    213 CGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTT 272

Qy    313 TTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCG 372
Db    273 TTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCG 332

Qy    373 GGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGG 432
Db    333 GGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGG 392

Qy    433 AGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCT 492
Db    393 AGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCCGGACTCGCCGCT 452

Qy    493 GGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAA 552
Db    453 GGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAA 512

Qy    553 CCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTT 612
Db    513 CCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTT 572

Qy    613 GGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGA 672
Db    573 GGGCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGA 632

Qy    673 GAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCC 732
Db    633 GAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCC 692

Qy    733 CATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGA-TCGGCAATCTGAAAGGCA 791
Db    693 CATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATTCGGCAATCTGAAAGGCA 752

Qy    792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
Db    753 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 812

Qy    852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCG-TTTCTGTGTCCTTTTTT 910
Db    813 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTTCTGTGTCCTTTTTT 872

Qy    911 ATTATTACGGC-GGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAA 969
Db    873 ATTATTACGGCGGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAA 932

Qy    970 TGCAAGAGCTCAAAGCAGGAAGCAGAGG-CAATTAAAGGCAGATGCTAAAAAAGCTATTG 1028
Db    933 TGCAAGAGCTCAAAGCAGGAAGCAGAGGCCAATTAAAGGCAGATGCTAAAAAAGCTATTG 992

Qy   1029 GAAGGC-TTCAACTACGCACACTGAAACAAGGAGACAA-GGAAATTGGCCCTGAT-GGAG 1085
Db    993 GAAGGCTTTCAACTACGCACACTGAAACAAGGAGACAAGGGAAATTGGCCCTGATGGGAG 1052

Qy   1086 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 1145
Db    1053 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 1112

Qy   1146 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCC 1205
Db    1113 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCC 1172

Qy   1206 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 1265
Db    1173 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 1232

Qy   1266 CAGTGT-CTTTACAAGTCCCTGT----ATCCAATGAAATATCTAATAGTGCCTCCTCCCA 1320
Db    1233 CAGTGTCCTTTACAAGTTTCCCTGATATCCAATAGAAATATCTAATAGTGCCTCCTCCCA 1292

Qy   1321 TGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGA 1380
Db    1293 TGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGA 1352

Qy   1381 TGAACCGCC-TCTGGAGGAACACGTGCAGTCAAC-AAATGAAAGTCTACAGCTGGT-AAA 1437
Db    1353 TGAACCGCCTTCTGGAGGAACACGTGCAGTCAACAAAATGAAAGTCTACAGCTGGTAAAA 1412

Qy   1438 CCATGAAGCAAATTCT-GTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACC---- 1492
Db    1413 CCATGAAGCAAATTCTGGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCCTTN 1472

Qy   1493 ---TTTGAAGAAGAC----GAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATC 1545
Db    1473 TTTTTGGAAGAAGACTGGAAAAANCTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATC 1532

Qy   1546 TTAAAATCTGTGTAAATAGAAAAC--TTGAACCATTAGTAATAACAGAACTGCCAATCAG 1603
Db    1533 TTAAAATCTGTGTAAATAGAAAACTGTGAACCATTAAGTAATAACAGAACTGCCAATCAG 1592

Qy   1604 GGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTG 1663
Db    1593 GGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTG 1652
```

```
Qy   1664 CTCAGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACT 1723
          ||||||||||||||||||||||||||||||||   |  | ||||||||||||||||||||
Db   1653 CTCAGATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACT 1710

Qy   1724 TTTTTCCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1711 TTTTTCCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1760
```

RESULT 5
AAA15985
ID AAA15985 standard; cDNA; 1284 BP.
XX
AC AAA15985;
XX
DT 12-JUN-2000 (first entry)
XX
DE Human protein clone HP10574 coding sequence.
XX
KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW coagulation disorder; myocardial infarction; inflammatory condition;
KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW nephritis; therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200005367-A2.
XX
PD 03-FEB-2000.
XX
PF 22-JUL-1999; 99WO-JP03929.
XX
PR 24-JUL-1998; 98JP-0208820.
PR 07-AUG-1998; 98JP-0224105.
PR 25-AUG-1998; 98JP-0238116.
PR 09-SEP-1998; 98JP-0254736.
PR 29-SEP-1998; 98JP-0275505.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2000-182694/16.
DR P-PSDB; AAY94897.
XX
PT Novel human proteins having hydrophobic domains useful for treating
PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS Claim 3; Page 330; 351pp; English.
XX
CC This sequence encodes a human protein of the invention, which has
CC hydrophobic domains. The DNA sequences can be used as a probe or as a
CC genetic marker. The protein can also be used as a marker, and to identify
CC potential genetic disorders. The DNA and protein can also be used as
CC nutritional sources or supplements. The protein exhibits cytokine, cell
CC proliferation, cell differentiation activities and induces production of
CC other cytokines in certain cell populations. The protein also exhibits
CC immune stimulating or immune suppressing activity. It can be used in the
CC treatment of various immune deficiencies and disorders, and to treat
CC infectious diseases caused by viral, bacterial, fungal or other
CC infections. The protein is also used for treating autoimmune disorders
CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC arthritis. It is also useful in the treatment of allergic reactions and
CC conditions such as asthma, and in immune suppression after organ
CC transplantation. The protein is useful in regulation of haematopoiesis
CC and consequently in the treatment of myeloid or lymphoid cell
CC deficiencies. It is also used in compositions for tissue growth or
CC regeneration. The protein is also used in the treatment of osteoporosis
CC or osteoarthritis and in the treatment of periodontal disease and other
CC tooth repair processes. The protein is used in the treatment of nervous
CC system disorders such as Alzheimer's disease, Parkinson's disease, and
CC Huntington's disease. They are useful for protection or regeneration and
CC treatment of lung or liver fibrosis, reperfusion injury in various
CC tissues, and conditions resulting from systemic cytokine damage. They are
CC also used for promoting or inhibiting tissue differentiation. They are
CC also used as contraceptives since they exhibit activin or inhibin related
CC activities and as a fertility inducing therapeutic. They are used for
CC treating various coagulation disorders and in treatment and prevention of
CC conditions resulting from coagulation activities e.g. myocardial
CC infarction or stroke. They also acts as receptors, receptor ligands or
CC inhibitors or agonists of receptor/ligand interactions. They are used to
CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC reperfusion injury, arthritis, and nephritis. They can be used to
CC prevent tumours.
XX
SQ Sequence 1284 BP; 323 A; 305 C; 352 G; 304 T; 0 other;

Query Match 72.1%; Score 1278.2; DB 21; Length 1284;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy    263 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 60

Qy    323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 120

Qy    383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 180

Qy    443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 240

Qy    503 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 300

Qy    563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 360

Qy    623 CAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    361 CAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGAGGGGCG 420

Qy    683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 480

Qy    743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 540

Qy    803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 600

Qy    863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 660

Qy    923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
```

```
Db     661 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 720

Qy     983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 780

Qy    1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 840

Qy    1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 900

Qy    1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
           |||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||
Db     901 AAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTGCAAATGTGAC 960

Qy    1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1020

Qy    1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1080

Qy    1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1140

Qy    1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1200

Qy    1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1201 GATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1260

Qy    1523 ACTGCTGTTCGAGAAATTAAATCT 1546
           |||||||||||||||||||||||
Db    1261 ACTGCTGTTCGAGAAATTAAATCT 1284
```

RESULT 6
AAS25884
ID AAS25884 standard; cDNA; 2306 BP.
XX
AC AAS25884;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 63.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR P-PSDB; AAU15897.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; SEQ ID No 63; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

Query Match 66.4%; Score 1178.8; DB 22; Length 2306;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1217; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

```
Qy    547 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 71

Qy    607 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 131

Qy    667 TTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 726
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 TTATGAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 191

Qy    727 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 251

Qy    787 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 311

Qy    847 AGGGAAAAAAACATGGCCCTTGGGTGAATCACTATTCAA-TTTTTTTCGTTTCTGTGTCCT 905
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    312 AGGGAAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTTCGTTTCTGTGTCCT 371

Qy    906 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 431

Qy    966 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 491

Qy   1026 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAG 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAG 551

Qy   1086 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 611

Qy   1146 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCC 1205
```

```
Db    612 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCC 671

Qy   1206 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 1265

Db    672 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 731

Qy   1266 CAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAG 1325

Db    732 CAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAG 791

Qy   1326 AGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAAC 1385

Db    792 AGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAAC 851

Qy   1386 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 1445

Db    852 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 911

Qy   1446 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACG 1505

Db    912 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACG 971

Qy   1506 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1565

Db    972 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1031

Qy   1566 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAAT 1625

Db   1032 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAAT 1091

Qy   1626 TGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCT 1685

Db   1092 TGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCT 1151

Qy   1686 ATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATA 1745

Db   1152 ATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATA 1208

Qy   1746 ATATTTTTCATAGGCAAGTTTCCTCTCAG 1774

Db   1209 ATATTTTTCATAGGCAAGTTTCCTCTCAG 1237
```

RESULT 7
ABL90796
ID ABL90796 standard; cDNA; 1249 BP.
XX
AC ABL90796;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1358.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
DR P-PSDB; ABB90387.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 4; SEQ ID NO 1358; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1249 BP; 379 A; 258 C; 273 G; 335 T; 4 other;

Query Match 65.8%; Score 1167.2; DB 24; Length 1249;
Best Local Similarity 98.7%; Pred. No. 2.2e-314;
Matches 1206; Conservative 1; Mismatches 11; Indels 4; Gaps 3;

```
Qy    547 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 606

Db     18 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 77

Qy    607 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 666

Db     78 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 137

Qy    667 TTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 726

Db    138 TTATGAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 197

Qy    727 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 786

Db    198 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 257

Qy    787 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 846

Db    258 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 317

Qy    847 AGGGAAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTT 906

Db    318 AGGGAAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTT 377

Qy    907 TTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACG 966

Db    378 TTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACG 437

Qy    967 GAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTAT 1026

Db    438 GAATGCAAGAGCTCAAAGCAGGAAGCANAGGCAATTAAAGGCAGATGCTAAAAAAGCTAT 497

Qy   1027 TGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGG-AAATTGGCCCTGATGGAG 1085

Db    498 TGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGKAAATTGGCCCTGATGGAG 557

Qy   1086 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 1145
```

```
Db    558 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 617

Qy   1146 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCC 1205
          ||||||||||||||||||||||||||||||||||:|||||||| |||||| ||||||||
Db    618 GCAACCATATTTTCCATAAGACATGTGTTGACCCMTGGCTGTTAGAACACAGGACTTGCC 677

Qy   1206 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 737

Qy   1266 CAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAG 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 CAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAG 797

Qy   1326 AGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAAC 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 AGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAAC 857

Qy   1386 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 917

Qy   1446 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACG 1505
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    918 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACG 977

Qy   1506 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1037

Qy   1566 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAAT 1625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAAT 1097

Qy   1626 TGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCT 1685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 TGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCT 1157

Qy   1686 ATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATA 1745
          |||||||||    |  | |||||||||||||||||||||||||||| ||||||||||||
Db    1158 ATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATA 1214

Qy   1746 ATATTTTTCATAGGCAAGTTTC 1767
          |||||||||||||||| || |
Db    1215 ATATTTTTCATAGGCAATTTCC 1236
```

RESULT 8
AAI72094
ID AAI72094 standard; cDNA; 2145 BP.
XX
AC AAI72094;
XX
DT 25-MAR-2002 (first entry)
XX
DE Mouse GRAIL cDNA.
XX
KW Murine; human; GRAIL; anergy; attenuation; tyrosine; phosphorylation;
KW antigenic stimulation; interleukin-2; gene therapy; polymorphism; IL-2;
KW autoimmune disease; tumour cell; cancer; transplant rejection; ss.
XX
OS Mus musculus.
XX
FH Key   Location/Qualifiers
FT CDS   358..1641
FT   /*tag= a
FT   /product= "GRAIL"
XX
PN WO200185943-A1.
XX
PD 15-NOV-2001.
XX
PF 11-MAY-2001; 2001WO-US15385.
XX
PR 11-MAY-2000; 2000US-203513P.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Ford GS, Bloom D, Fathman CG;
XX
DR WPI; 2002-055597/07.
DR P-PSDB; AAB47799.
XX
PT Novel nucleic acid sequences of anergy associated genes, including
PT GRAIL gene useful in the evaluation of pathophysiology or immunotherapy
PT of cancer, autoimmune disease and transplant rejection -
XX
PS Claim 2; Page 42-44; 50pp; English.
XX
CC This sequence encodes murine GRAIL protein. The GRAIL gene is an
CC anergy associated gene which is upregulated during the early stages of
CC induction of anergy. GRAIL has been shown to attenuate IL-2
CC transcription in T-cells during response to antigenic stimulation.
CC GRAIL DNA is useful for decreasing the responsiveness of a T cell
CC population, especially synthesis of interleukin-2 (IL-2) in response to
CC antigenic stimulation, by up-regulating GRAIL activity in the T cell
CC population. GRAIL DNA is useful in producing compositions that modulate
CC induction or maintenance of anergy, for gene therapy, mapping
CC functional regions of the encoded protein, to analyse a patient sample
CC for the presence of polymorphisms or alterations in expression of
CC sequences associated with T cell anergy, disease states, genetic
CC predisposition to a disease state, and in studying associated
CC physiological pathways. Modulation of the gene activity in vivo is
CC useful for prophylactic and therapeutic purposes, such as treating
CC autoimmune disease and to enhance immune response to tumour cells, and
CC identification of anergic T cells. GRAIL DNA is useful in the
CC evaluation of the pathophysiology or immunotherapy of cancer,
CC autoimmune disease, and transplant rejection. Genetic sequences
CC involved in anergy induction are useful as markers in the evaluation of
CC specific immunotherapies. Functional characterization of genes involved
CC in anergy induction allows the elucidation of the mechanisms of T cell
CC anergy, including the transcriptional blockade of IL-2, which may be
CC manipulated to regulate T cell responses in human disease. Modulation
CC of expression of GRAIL gene is useful in manipulating the anergic
CC state. The genetic sequences find use alone or in combinations in
CC determining the expression profile of cells relating to anergy, for
CC example in screening of candidate biologically active compounds for
CC modulation of T cell anergy. Expression of GRAIL has been found to
CC block tyrosine phosphorylation of a protein present in T cells during
CC activation.
XX
SQ Sequence 2145 BP; 525 A; 523 C; 560 G; 537 T; 0 other;

Query Match 65.7%; Score 1166; DB 24; Length 2145;
Best Local Similarity 89.2%; Pred. No. 5.4e-314;
Matches 1359; Conservative 0; Mismatches 151; Indels 13; Gaps 9;

```
Qy    252 GCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCT 311
          | | |||||||||||||||||||||| |||  |||||||| |||||||| ||||||||
Db    347 GAGTGCGCGCCATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAG 406

Qy    312 TTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCC 371
          | || |  | ||||| |||||||||| ||||| |||||||||| || ||||||||||
Db    407 CTGCCCGGCTACTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCC 466

Qy    372 GGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGG 431
          |||| || ||||| |||||||| |||||||||||||||||||||||||||||||||| |
Db    467 GGGGAGCCGAAGCCGTGTGGACTGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCG 526

Qy    432 GAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 491
          ||||| |||| ||||||| |||||||||||| ||||||||||||||||||||||||||||
Db    527 GAGTG-ACCGCACGGTGT-GGAGCTGAGCGA-GAGGGCGTGTACGGCCAGGACTCGCCGC 583

Qy    492 TGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTA 551
          || |||| ||  | |||||||||||||||||||||||||||||||||| ||||||||||
```

```
Db    584 TGAAGCCCGTCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTA 643

Qy    552 ACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGT 611
          |||||||||| |||||||||||||||||||||||||||||| ||||| |||||||| |||||||
Db    644 ACCCGCACACCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGT 703

Qy    612 TGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATG 671
          |||||||||||||||||| || |||||||||||||| |||||||||||||||||||   |
Db    704 TGGCCCTCATCCAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAG 763

Qy    672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
          |||||  |||| |||||||| |||||||||||||||| |||||||||||||||||||||
Db    764 AGAGAGGGGCTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCC 823

Qy    732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
          |||||||||||||||||| | ||||||||||||||||||| ||||||||||||| |
Db    824 CCATGTCTCACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAA 883

Qy    792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
          |||||||||||||  |||||||||||||||||| ||||| |||||||||||  ||||||||||
Db    884 CAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGA 943

Qy    852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTA 911
          |||||||||||||||||||||||||||| |||||||||| |||||||||||||||||||| |
Db    944 AAAAACATGGCCCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCA 1003

Qy    912 TTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATG 971
          | |||||||| ||||| |||||||||| |||||||||||||||||||  |||| ||||
Db   1004 TAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATG 1063

Qy    972 CAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAA 1031
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1064 CAAGAGCTCAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAA 1123

Qy   1032 GGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTT 1091
            |||||| || |||||  |||||||||||||||||||||||||||||||||||||||| |
Db   1124 AGCTTCAGCTGCGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCT 1183

Qy   1092 GTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACC 1151
          |||||||||||||||  | ||||| |||||||||||||| |||||| |||| ||||| |
Db   1184 GTGCTGTGTGCATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATC 1243

Qy   1152 ATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGT 1211
          |||||||||||||||||||||| ||||| ||||| ||| |||||| ||||||||||||||
Db   1244 ATATTTTCCATAAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGT 1303

Qy   1212 GCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1271
          |||| |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db   1304 GCAAGTGTGACATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1363

Qy   1272 CTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATA 1331
          |||||||||| ||||| || |||||| |||||||| |||||| |||||||||||||| |
Db   1364 CTTTACAAGTTCCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACA 1423

Qy   1332 ATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTC 1391
            |||||| |||||  |||||||||||||||||||||||||||| ||| |||||||  || ||||
Db   1424 GTCGCAGTGAGACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTC 1483

Qy   1392 TGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATT 1451
          |||||||||| | ||||||| ||||||||| |||||||||||||||||||||||||||||
Db   1484 TGGAGGAACATGCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATT 1543

Qy   1452 CTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTC 1511
          ||||||| ||||||||| |||| ||||| |||||||||||||||||||| ||||||||
Db   1544 CTGTGGCCGTGGATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTC 1603

Qy   1512 CTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTT-AAAATCTGTGTAAATAGAAAACT 1570
          || |||||||| | |||||||| || |||||||||| |||||||||| ||||||||||||
Db   1604 CTGATCAAGAGGCAGCTGTTCGGGAGATTAAATCTTAAAAATCTGTGTCAATAGAAAACT 1663

Qy   1571 TGAACCATTAG-TAATAACAGAACTGCCAATCAGGGCCTAGTTT-CTATTAATAAATTGG 1628
          |||||| |||| ||| ||||| |||||||||||||||||||| |||| ||| || |||
Db   1664 TGAACCGTTAGTTAACAACAGGACTGCCAATCAGGGCCTAGTTTACTATGAATGAACTGG 1723

Qy   1629 ATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATA 1688
            ||||     ||||| |||| |||||||||||||||| || | ||| ||||||| |||||
Db   1724 GTAAA---CGTAAAACAAGAATGATACTGAAAGTGCTGAGGTAACTTATATTATACTATA 1780

Qy   1689 GTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATA 1748
          ||||||   |  | || ||||||| ||   || | ||||  |  |||||||  ||||
Db   1781 GTTAAA--TGGCTTAACATATTTACCCCAGTACCGTTTT--CCACAAACTCACCATAACG 1836

Qy   1749 TTTTTCATAGGCAAGTTTCCTCT 1771
          |||||||||||||||||||||||
Db   1837 TTTTTCATAGGCAAGTTTCCTCT 1859
```

RESULT 9
AAS26340
ID AAS26340 standard; cDNA; 1250 BP.
XX
AC AAS26340;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 519.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR P-PSDB; AAU16353.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; SEQ ID No 519; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

Query Match 65.3%; Score 1157.6; DB 22; Length 1250;
Best Local Similarity 98.8%; Pred. No. 8.9e-312;

```
  Matches 1208;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    4;

Qy     547 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC  77

Qy     607 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 137

Qy     667 TTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 726
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 TTATGAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 197

Qy     727 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 257

Qy     787 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 317

Qy     847 AGGGAAAAAAACATGGCCCTTGGGTGAATCACTATTCAA-TTTTTTTCGTTTCTGTGTCCT 905
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     318 AGGGAAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTTCGTTTCTGTGTCCT 377

Qy     906 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     378 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 437

Qy     966 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 1025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     438 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 497

Qy    1026 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGG-AAATTGGCCCTGATGGA 1084
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     498 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGNAAATTGGCCCTGATGGA 557

Qy    1085 GATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACG 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     558 GATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACG 617

Qy    1145 TGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGC 1204
           |||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db     618 TGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGC 677

Qy    1205 CCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGA 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 CCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGA 737

Qy    1265 TCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAA 1324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 TCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAA 797

Qy    1325 GAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAA 1384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 GAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAA 857

Qy    1385 CCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAA 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 CCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAA 917

Qy    1445 GCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGAC 1504
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     918 GCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGAC 977

Qy    1505 GAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAG 1564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 GAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAG 1037

Qy    1565 AAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAA 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 AAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAA 1097

Qy    1625 TTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGC 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 TTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGC 1157

Qy    1685 TATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTAT 1744
           |||||||||| | | ||||||||||||||||||||||||||||| ||||||||||||
Db    1158 TATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTAT 1214

Qy    1745 AATATTTTTCATAGGCAAGTTTC 1767
           |||||||||||||||||| || |
Db    1215 AATATTTTTCATAGGCAATTTCC 1237
```

RESULT 10
ABK12983
ID ABK12983 standard; cDNA; 1287 BP.
XX
AC ABK12983;
XX
DT 23-APR-2002 (first entry)
XX
DE Mouse goliath cDNA sequence.
XX
KW Mouse; goliath protein; antiangiogenic; vasotropic; gene therapy;
KW dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW unrecovered nerve trauma; gene; ss.
XX
OS Mus sp.
XX
FH Key   Location/Qualifiers
FT CDS   1..1287
FT   /*tag= a
FT   /product= "Mouse_goliath_protein"
XX
PN WO200193681-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-US18000.
XX
PR 02-JUN-2000; 2000US-0586398.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Harland R, Baker JC;
XX
DR WPI; 2002-147637/19.
DR P-PSDB; AAU74918.
XX
PT New compositions comprising goliath proteins, useful for modulating
PT angiogenesis or neurogenesis in mammals e.g. for preventing or
PT treating undesirable vascularisation of a tumour, ischaemia or
PT neurodegenerative disease -
XX
PS Examples; Page 38; 45pp; English.
XX
CC The present invention relates to a new pharmaceutical composition that
CC comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC has a sequence identity of at least 75% to the protein sequences
CC (AAU74918-AAU74921) fully defined in the specification. The composition
CC is useful for modulating angiogenesis or neurogenesis in mammals,
CC particularly in humans or mice. Specifically, the composition is useful
CC for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC The present nucleic acid sequence encodes the mouse goliath protein
CC that was used in the invention for modulating angiogenesis or
CC neurogenesis.
XX
SQ Sequence 1287 BP; 308 A; 320 C; 359 G; 300 T; 0 other;

Query Match 62.4%; Score 1106.8; DB 24; Length 1287;
Best Local Similarity 91.2%; Pred. No. 1.3e-297;
Matches 1174; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```
Qy    263 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
Db      1 ATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAGCTGCCCGGCTA 60

Qy    323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
Db     61 CTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCCGGGGAGCCGAA 120

Qy    383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
Db    121 GCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCGGAGTGAACCGC 180

Qy    443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
Db    181 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCCGTC 240

Qy    503 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
Db    241 TCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTAACCCGCACACC 300

Qy    563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
Db    301 AATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGTTGGCCCTCATC 360

Qy    623 CAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
Db    361 CAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAGAGAGAGGGGCT 420

Qy    683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
Db    421 TCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 480

Qy    743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
Db    481 CCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAACAAAAATTCTG 540

Qy    803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
Db    541 CAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGAAAAAACATGGC 600

Qy    863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
Db    601 CCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCATAATTACGGCA 660

Qy    923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
Db    661 GCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATGCAAGAGCTCAA 720

Qy    983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
Db    721 AGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAAAGCTTCAGCTG 780

Qy   1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
Db    781 CGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCTGTGCTGTGTGC 840

Qy   1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
Db    841 ATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATCATATTTTCCAT 900

Qy   1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
Db    901 AAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGTGCAAGTGTGAC 960

Qy   1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
Db    961 ATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTT 1020

Qy   1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1342
Db   1021 CCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACAGTCGCAGTGAG 1080

Qy   1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
Db   1081 ACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTCTGGAGGAACAT 1140

Qy   1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
Db   1141 GCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCCGTG 1200

Qy   1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
Db   1201 GATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTCCTGATCAAGAG 1260

Qy   1523 ACTGCTGTTCGAGAAATTAAATCTTAA 1549
Db   1261 GCAGCTGTTCGGGAGATTAAATCTTAA 1287
```

RESULT 11
ABK12992
ID ABK12992 standard; DNA; 1287 BP.
XX
AC ABK12992;
XX
DT 23-APR-2002 (first entry)
XX
DE Synthetic goliath DNA sequence #6.
XX
KW Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW unrecovered nerve trauma; ds.
XX
OS Synthetic.
XX
PN WO200193681-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-US18000.
XX
PR 02-JUN-2000; 2000US-0586398.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Harland R, Baker JC;
XX
DR WPI; 2002-147637/19.
XX
PT New compositions comprising goliath proteins, useful for modulating
PT angiogenesis or neurogenesis in mammals e.g. for preventing or
PT treating undesirable vascularisation of a tumour, ischaemia or
PT neurodegenerative disease -
XX
PS Disclosure; Page 42; 45pp; English.
XX
CC The present invention relates to a new pharmaceutical composition that
CC comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC has a sequence identity of at least 75% to the protein sequences
CC (AAU74918-AAU74921) fully defined in the specification. The composition
CC is useful for modulating angiogenesis or neurogenesis in mammals,
CC particularly in humans or mice. Specifically, the composition is useful
CC for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC The present nucleic acid sequence represents synthetic DNA sequence #6
CC that is one of several artificial goliath nucleic acids (ABK12987-
CC ABK12992) used in the invention for modulating angiogenesis or
CC neurogenesis.

```
XX
SQ   Sequence 1287 BP; 312 A; 321 C; 357 G; 297 T; 0 other;

  Query Match             59.3%;  Score 1052.4;  DB 24;  Length 1287;
  Best Local Similarity   88.6%;  Pred. No. 1.8e-282;
  Matches 1140;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    263 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
Db      1 ATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGGGGCTGCGGAGCTGCCCGACTA 60

Qy    323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
Db     61 CTGGCTTGGTGCTTTCTTCTGGCTCTGAGTCCCCACGCGCCCGGTTCCCGGGGAGCCGAA 120

Qy    383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
Db    121 GCCGTGTGAACCGCGTACCTCAACGTTTCCTGGCGGGTTCCGCACACCGGAGTGAACCGC 180

Qy    443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
Db    181 ACGGTGTGGGAGCTGAGCGAAGAGGGCGTGTACGGCCATGACTCGCCGCTGGAGCCCGTC 240

Qy    503 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
Db    241 TCCGGGGTCCTGGTGCCGCCCGACGGGCCCGGAGCGCTCAACGCCTGTAATCCGCACACC 300

Qy    563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
Db    301 AATTTCACCGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCATGGTTGGCCCTCATC 360

Qy    623 CAACGCGGCGGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
Db    361 CATCGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCAGCTGGCTTCAGAGAGAGGAGCT 420

Qy    683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
Db    421 TCTGGAGCGGTCATTTTTAACTTCCCTGGGACCCGCAATGAGGTCATCCCGATGTCTCAC 480

Qy    743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
Db    481 CCGGGTGCAGGGGACATTGTTGCAATTATGATTGGCAATCTGAACGGAACAAAAATTCTG 540

Qy    803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
Db    541 CAGTCTATTCAAAGAGGCATACAAGTCACAATGGTCATTGAAGTAGGGAAAAAACACGGC 600

Qy    863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
Db    601 CCTTGGGTGAATCAGTATTCAATTTTCTTCGTATCTGTGTCCTTTTTTCATTATTACGGCA 660

Qy    923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
Db    661 GCAACCGTCGGCTATTTCATCTTTTAGTCTGCTCGAAGATTACGAAATGCAAGAGCTCAA 720

Qy    983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
Db    721 AGTAGGAAGCAGAGGCAGTTCAAGGCAGATGCTAAAAAGGCTATTGGAAAGCTTCAACTG 780

Qy   1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
Db    781 CGCACCTTGAAACATGGAGACAAGGAAATTGGCCCTGATGGAGATAGCTGGGCTGTGTGC 840

Qy   1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
Db    841 ATTGAGCTATATAAGCCAAATGATTTTGTGCGCATCCTAACCTGCAATCATATTTTCCAT 900

Qy   1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
Db    901 AAGACATGTGTGGACCCGTGACTTTTAGAACACAGGACTTGCCCCATGTGCAAGTGCGAC 960

Qy   1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
Db    961 ATTCTCAAAGCTCTGGGAATTGAGGTGGATGTAGAAGATGGATCAGTGTCTTTACAAGTT 1020

Qy   1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1342
Db    1021 CCTGTTTCCAATGAAGCATCTAATACGGCCTCTCCCCATGAAGAAGACAGTCGCAGTGAG 1080

Qy   1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
Db    1081 ACTGCATCATCTGGATATGCCTCAGTACAAGGAGCAGAGGAGCCACCTCTGGAGGAACAT 1140

Qy   1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
Db    1141 GCGCAGTCAGCAAATGAAAATCTACAGCTGGTCAACCATGAAGCAAATTCGGTGGCCGTG 1200

Qy   1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
Db    1201 GATGTTGTACCCCATGTTGACAACCCTACCTTTGAAGAAGATGACACTCCTGATCAAGAG 1260

Qy   1523 ACTGCTGTTCGAGAAATTAAATCTTAA 1549
Db    1261 GCGGCTGTTCGGGAGATTAAATCTTAA 1287


RESULT 12
ABK12990
ID   ABK12990 standard; DNA; 1287 BP.
XX
AC   ABK12990;
XX
DT   23-APR-2002  (first entry)
XX
DE   Synthetic goliath DNA sequence #4.
XX
KW   Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW   dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW   cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW   unrecovered nerve trauma; ds.
XX
OS   Synthetic.
XX
PN   WO200193681-A1.
XX
PD   13-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-US18000.
XX
PR   02-JUN-2000; 2000US-0586398.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Harland R,  Baker JC;
XX
DR   WPI; 2002-147637/19.
XX
PT   New compositions comprising goliath proteins, useful for modulating
PT   angiogenesis or neurogenesis in mammals e.g. for preventing or
PT   treating undesirable vascularisation of a tumour, ischaemia or
PT   neurodegenerative disease  -
XX
PS   Disclosure; Page 41; 45pp; English.
XX
CC   The present invention relates to a new pharmaceutical composition that
CC   comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC   The present nucleic acid sequence represents synthetic DNA sequence #4
CC   that is one of several artificial goliath nucleic acids (ABK12987-
```

CC ABK12992) used in the invention for modulating angiogenesis or
CC neurogenesis.
XX
SQ Sequence 1287 BP; 312 A; 321 C; 356 G; 298 T; 0 other;

Query Match 58.8%; Score 1042.8; DB 24; Length 1287;
Best Local Similarity 88.1%; Pred. No. 8.7e-280;
Matches 1134; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy    263 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
          ||||||||||||||| |||    ||||||| ||||||||||||||||||    | || |  |
Db      1 ATGGGGCCGCCGCCCGGGATAGGGGTCTACTGCCGCGGTGGCTGCGGAGCTGCCCGCCTA 60

Qy    323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
          |||||  |||||||  || ||||| ||||||||  ||  || ||||||||||  ||  ||  |||
Db     61 CTGGCTTGGTGCTTGCTTCTGGCTCTGAGTCCACACGCGCCCGGTTCCCGTGGAGCCGAA 120

Qy    383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
          ||  |||||  ||||||||||||||||||||||||||||||||||||||  ||  ||||||||||
Db    121 GCCGTGTGCACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCAAACCGGAGTGAACCGC 180

Qy    443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
          ||  ||||||||||||||||  ||||||||||||||||||||||||||||||||||||  ||
Db    181 ACTGTGTGGGAGCTGAGCGACGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCAGTC 240

Qy    503 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
          | ||||||||||  |||||||||||||||||  |||||  ||||||||||  ||||||||
Db    241 TCCGGGGTCCTGGTTCCGCCCGACGGGCCCGGCGCGCTCAACGCCTGTAAGCCGCACACC 300

Qy    563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
          ||||||||  ||||||||||||||||||  |||||  ||||||||  ||  ||||||||||||||||
Db    301 AATTTCACAGTGCCCACGGTTTGGGGTAGCACGGTGCAAGTATCCTGGTTGGCCCTCATC 360

Qy    623 CAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
          ||  |||||  ||  ||||||||  |||||  ||||||||||||||||||||  ||||||  |  ||
Db    361 CAGCGCGGTGGAGGCTGCACATTCGCGGACAAGATCCATCTGGCTTCAGAGAGAGGCGCT 420

Qy    683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
          ||||||||  |||||  ||||||||||||  |||||  ||||||||||||||||||  |||||||||
Db    421 TCTGGAGCGGTCATGTTTAACTTCCCTGGGACACGCAATGAGGTCATCCCTATGTCTCAC 480

Qy    743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
          ||||||||  |  ||||||||||||||  |||||  ||||||||||||||  ||||||||||||
Db    481 CCGGGTGCCGGGGACATTGTTGCAATGATGATTGGCAATCTGAAAGGAACAAAAATTCTG 540

Qy    803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
          ||  ||||||||||||||||  |||||  ||||||||||  ||||||||||||||||||  |||
Db    541 CATTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATGGAAGTAGGGAAAAAACAAGGC 600

Qy    863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
          ||||||||||||||  ||||||||||  |||||  ||||||||||||||  ||  |||||||||
Db    601 CCTTGGGTGAATCATTATTCAATTTTCTTCGTCTCTGTGTCCTTTTTTCATGATTACGGCA 660

Qy    923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
          |||||  ||  ||||||||  ||||||||||||||||||||  ||||  ||||||||||||||||
Db    661 GCAACCGTAGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGCAATGCAAGAGCTCAA 720

Qy    983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
          ||  ||||||||||||||  ||||||||||||||||||||||  ||||||||||  ||||||  ||
Db    721 AGGAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAATGCTATTGGAAAGCTTCACCTG 780

Qy   1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
          |||||  |||||||  ||||||||||||||||||  ||||||||||||||  ||||||||||||
Db    781 CGCACCTTGAAACAGGGAGACAAGGAAATTGGACCTGATGGAGATAGCTGTGCTGTGTGC 840

Qy   1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
          |||||  |  |||||  ||||||||||||||  ||||||  ||||  ||  ||  ||||||||||||
Db    841 ATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGAAATCATATTTTCCAT 900

Qy   1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
          ||  ||||||||  |||||  ||  ||  |||  ||||||  |||  ||||||||||||||  ||  |||
Db    901 AATACATGTGTGGACCCGTGCCTTTTAGAACACAGGACGTGCCCCATGTGCAAGTGAGAC 960

Qy   1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
          ||  |||||||||  |  ||||||||||||||||||  ||||||||||||||||||  |||||||
Db     961 ATTCTCAAAGCTCTTGGAATTGAGGTGGATGTCGAAGATGGATCAGTGTCGTTACAAGTT 1020

Qy   1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1342
          |||||  ||  ||||||  |||||||  ||||||  ||||||||||  ||  |  ||||||  |||
Db    1021 CCTGTTTCAAATGAAGCATCTAATACTGCCTCTCCCCATGAAGACGACAGTCGCAGTGAG 1080

Qy   1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
          ||  ||||||||||||||||||  ||||||||  |||  |||||||  ||  ||||||||||  ||
Db    1081 ACGGCATCATCTGGATATGCATCAGTACAAGGAGCAGATGAGCCACCTCTGGAGGACCAT 1140

Qy   1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
          |  |||||||  ||||  ||||  ||||||||||||||||||||||||||||||||||  |||
Db    1141 GCGCAGTCAGCAAAGGAAAATCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCCGTG 1200

Qy   1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
          ||||||  |  ||  |||||  ||||||||  ||||||||||||||  |||||||||  ||||||||
Db    1201 GATGTTGTCCCCCATGTTGACAACCCGACCTTTGAAGAAGATGAAACTCCTGATCAAGAG 1260

Qy   1523 ACTGCTGTTCGAGAAATTAAATCTTAA 1549
          ||||||||||  ||  |||||  ||||||
Db   1261 GCTGCTGTTCGGGAGATTAACTCTTAA 1287
```

RESULT 13
ABK12991
ID ABK12991 standard; DNA; 1287 BP.
XX
AC ABK12991;
XX
DT 23-APR-2002 (first entry)
XX
DE Synthetic goliath DNA sequence #5.
XX
KW Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW unrecovered nerve trauma; ds.
XX
OS Synthetic.
XX
PN WO200193681-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-US18000.
XX
PR 02-JUN-2000; 2000US-0586398.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Harland R, Baker JC;
XX
DR WPI; 2002-147637/19.
XX
PT New compositions comprising goliath proteins, useful for modulating
PT angiogenesis or neurogenesis in mammals e.g. for preventing or
PT treating undesirable vascularisation of a tumour, ischaemia or
PT neurodegenerative disease -
XX
PS Disclosure; Page 41-42; 45pp; English.
XX
CC The present invention relates to a new pharmaceutical composition that
CC comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC has a sequence identity of at least 75% to the protein sequences
CC (AAU74918-AAU74921) fully defined in the specification. The composition
CC is useful for modulating angiogenesis or neurogenesis in mammals,
CC particularly in humans or mice. Specifically, the composition is useful
CC for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.

```
CC   The present nucleic acid sequence represents synthetic DNA sequence #5
CC   that is one of several artificial goliath nucleic acids (ABK12987-
CC   ABK12992) used in the invention for modulating angiogenesis or
CC   neurogenesis.
XX
SQ   Sequence 1287 BP; 312 A; 321 C; 356 G; 298 T; 0 other;

  Query Match             58.8%;  Score 1042.8;  DB 24;  Length 1287;
  Best Local Similarity   88.1%;  Pred. No. 8.7e-280;
  Matches 1134;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy     263 ATGGGGCCGCCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
           ||||||||||||||| |||   ||||||| |||||||| ||||||||   | || || |
Db       1 ATGGGGCCGCCGCCCGGGATTGGGGTCTACTGCCGCGGCGGCTGCGGAGCTGCCCGACTA 60

Qy     323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
           |||||  |||||||| || |||||  |||||||| || || |||||||||| || || |||
Db      61 CTGGCTTGGTGCTTGCTTCTGGCTCTGAGTCCTCACGCGCCCGGTTCCCGCGGAGCCGAA 120

Qy     383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
           ||  |||||  ||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db     121 GCCGTGTGAACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCATACCGGAGTGAACCGC 180

Qy     443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
           ||  ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db     181 ACCGTGTGGGAGCTGAGCGAAGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTC 240

Qy     503 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
           | ||||||||||| |||||||||||||||| ||||| |||||||||| ||||||||
Db     241 TCCGGGGTCCTGGTCCCGCCCGACGGGCCCGGAGCGCTCAACGCCTGTAAGCCGCACACC 300

Qy     563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
           ||||||||  |||||||||||||||||||| |||||  |||||||| || ||||||||||||||
Db     301 AATTTCACTGTGCCCACGGTTTGGGGCAGCACGGTGCAAGTATCATGGTTGGCCCTCATC 360

Qy     623 CAACGCGGCGGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
           ||  |||||  || ||||||||  |||||  ||||||||||  |||||||   ||||||  | ||
Db     361 CAGCGCGGTGGAGGCTGCACTTTCGCGGACAAGATCCACCTGGCTTCAGAGAGAGGAGCT 420

Qy     683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
           ||||||||  |||||  ||||||||||  |||||  ||||||||||||||||||||||||||
Db     421 TCTGGAGCGGTCATGTTTAACTTCCCTGGGACTCGCAATGAGGTCATCCCCATGTCTCAC 480

Qy     743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
           ||||||||||   |||||||||||||| |||||  |||||||||| || ||||||||||||
Db     481 CCGGGTGCAGGGGACATTGTTGCAATGATGATTGGCAATCTGAATGGAACAAAAATTCTG 540

Qy     803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
           ||  ||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db     541 CACTCTATTCAAAGAGGCATACAAGTCACAATGGTCATGGAAGTAGGGAAAAAACATGGC 600

Qy     863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
           ||||||||||||||||||||||||||||  |||||  |||||||||||||| ||  ||||||||
Db     601 CCTTGGGTGAATCACTATTCAATTTTCTTCGTATCTGTGTCCTTTTTTCATGATTACGGCA 660

Qy     923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
           |||||  ||  ||||||||  ||||||||  ||||||||||   ||||  ||||||||||||||
Db     661 GCAACCGTTGGCTATTTCATCTTTTACTCTGCTCGAAGATTACGAAATGCAAGAGCTCAA 720

Qy     983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
           ||  |||||||||||||  ||  |||||||||||||||||  |||||||||  |||||||||
Db     721 AGGAGGAAGCAGAGGCAGTTTAAGGCAGATGCTAAAAACGCTATTGGAAAGCTTCAACTG 780

Qy    1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
           |||||  |||||||  ||||||||||||||||||  ||||||||||||||  ||  |||||||||
Db     781 CGCACCTTGAAACAGGGAGACAAGGAAATTGGTCCTGATGGAGATAGCTGCGCTGTGTGC 840

Qy    1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
           |||||  | |||||  ||||||||||||||  ||||||  ||||  ||  ||  ||||||||||||
Db     841 ATTGAGCTATATAAGCCAAATGATTTGGTGCGCATCCTAACCTGTAATCATATTTTCCAT 900

Qy    1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
           ||  ||||||||  |||||  ||  ||  |||  ||||||  |||  ||||||||||||||  ||||||

Db     901 AACACATGTGTGGACCCGTGACTTTTAGAACACAGGACGTGCCCCATGTGCAAGTGTGAC 960

Qy    1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
           ||  |||||||||  | ||||||||||||||||||  ||||||||||||||||||  |||||||
Db     961 ATTCTCAAAGCTCTCGGAATTGAGGTGGATGTAGAAGATGGATCAGTGTCGTTACAAGTT 1020

Qy    1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCCATGAAGAGGATAATCGCAGCGAG 1342
           |||||  ||  ||||||  ||||||||  |||||  ||||||||||  ||  |  ||||||  |||
Db    1021 CCTGTTTCTAATGAAGCATCTAATACCGCCTCTCCCCATGAAGAAGACAGTCGCAGTGAG 1080

Qy    1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
           ||  ||||||||||||||||||||||||||||  |||  ||||  ||  ||  ||||||||||||||
Db    1081 ACGGCATCATCTGGATATGCTTCAGTACAAGGAGCAGACGAGCCACCTCTGGAGGAACAT 1140

Qy    1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
           | |||||||  ||||  ||||  ||||||||||||  ||||||||||||||||||  |||||  |||
Db    1141 GCGCAGTCAGCAAAGGAAAATCTACAGCTGGTTAACCATGAAGCAAATTCCGTGGCCGTG 1200

Qy    1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
           ||||||  |  ||  |||||  ||||||||  ||||||||||||||  ||  ||||||  ||||||||
Db    1201 GATGTTGTACCCCATGTTGACAACCCGACCTTTGAAGAAGATGATACTCCTGATCAAGAG 1260

Qy    1523 ACTGCTGTTCGAGAAATTAAATCTTAA 1549
           |  ||||||||  ||  ||||||||||||
Db    1261 GCCGCTGTTCGGGAGATTAAATCTTAA 1287


RESULT 14
AAI59292
ID   AAI59292 standard; cDNA; 1135 BP.
XX
AC   AAI59292;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1495.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40136.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
```

```
PS   Claim 1; SEQ ID NO 1495; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1135 BP; 367 A; 218 C; 234 G; 316 T; 0 other;

  Query Match             57.5%;  Score 1019.4;  DB 22;  Length 1135;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-273;
  Matches 1057;  Conservative    0;  Mismatches    7;  Indels    4;  Gaps    3;

Qy     708 CCGGGACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTGCA-GTAGACATTGTTGCA 766
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      26 CCGGGACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTGCATGTAGACATTGTTGCA 85

Qy     767 ATCATGATCGGCAATCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAA 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 ATCATGATCGGCAATCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAA 145

Qy     827 GTGACAATGGTCATAGAAGTAGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATT 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     146 GTGACAATGGTCATAGAAGTAGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATT 205

Qy     887 TTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTT 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206 TTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTT 265

Qy     947 TATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAG 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     266 TATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAG 325

Qy    1007 GCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAG 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326 GCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAG 385

Qy    1067 GAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGAT 1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     386 GAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGAT 445

Qy    1127 TTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTG 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     446 TTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTG 505

Qy    1187 TTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAG 1246
           ||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     506 TTAGAACACAGGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAG 565

Qy    1247 GTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAAT 1306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     566 GTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAAT 625

Qy    1307 AGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCA 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     626 AGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCA 685

Qy    1367 GTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTA 1426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     686 GTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTA 745

Qy    1427 CAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAAC 1486
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     746 CAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAAC 805

Qy    1487 CCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCT 1546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     806 CCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCT 865

Qy    1547 TAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGC 1606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     866 TAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGC 925

Qy    1607 CTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTC 1666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     926 CTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTC 985

Qy    1667 AGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTT 1726
           |||||||||||||||||||||||||||||    |  | ||||||||||||||||||||||
Db     986 AGATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTT 1043

Qy    1727 TTCCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
           ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1044 TTCCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1090
```

```
RESULT 15
ABK12989
ID   ABK12989 standard; DNA; 1287 BP.
XX
AC   ABK12989;
XX
DT   23-APR-2002  (first entry)
XX
DE   Synthetic goliath DNA sequence #3.
XX
KW   Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW   dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW   cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW   unrecovered nerve trauma; ds.
XX
OS   Synthetic.
XX
PN   WO200193681-A1.
XX
PD   13-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-US18000.
XX
PR   02-JUN-2000; 2000US-0586398.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Harland R,  Baker JC;
XX
DR   WPI; 2002-147637/19.
XX
PT   New compositions comprising goliath proteins, useful for modulating
PT   angiogenesis or neurogenesis in mammals e.g. for preventing or
PT   treating undesirable vascularisation of a tumour, ischaemia or
PT   neurodegenerative disease  -
XX
PS   Disclosure; Page 40-41; 45pp; English.
XX
CC   The present invention relates to a new pharmaceutical composition that
CC   comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
```

CC The present nucleic acid sequence represents synthetic DNA sequence #3
CC that is one of several artificial goliath nucleic acids (ABK12987-
CC ABK12992) used in the invention for modulating angiogenesis or
CC neurogenesis.
XX
SQ Sequence 1287 BP; 308 A; 320 C; 355 G; 304 T; 0 other;

Query Match 56.3%; Score 999.6; DB 24; Length 1287;
Best Local Similarity 86.0%; Pred. No. 9.5e-268;
Matches 1107; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```
Qy    263 ATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTTTCCAGATTG 322
          ||||||||||| || |||   |||||||  ||||||||||||||||||  | || |  |
Db      1 ATGGGGCCGCCCCCCGGGATGGGGGTCTAATGCCGCGGTGGCTGCGGCGCTGCCCGGCTA 60

Qy    323 CTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAA 382
          ||||||||||||||| || ||||||||||||||||| || || |||||||| || || ||
Db     61 CTGGCATGGTGCTTTCTTCTGGCCCTGAGTCCGCACGCGCCAGGTTCCCGTGGAGCCGAC 120

Qy    383 GCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGT 442
          || |||||||||||||| |||||||| |||||||| || ||||| |||||
Db    121 GCCGTGTGGACCGCGTAACTCAACGTTTCCTGGCGCGTTCCGCAGACCGGAGTAAACCGC 180

Qy    443 ACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTG 502
          || |||||||| |||||||||||||||| |||||||| |||||||| |||||||| ||
Db    181 ACTGTGTGGGACCTGAGCGAGGAGGGCGTATACGGCCATGACTCGCCCCTGGAGCCGGTC 240

Qy    503 GCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACG 562
          | || |||||||| |||||||||||||||||||||| || |||||||| ||||||||
Db    241 TCCGGAGTCCTGGTTCCGCCCGACGGGCCCGGGGCGCTCAAAGCCTGTAATCCGCACACC 300

Qy    563 AATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATC 622
          |||||||||||||||| |||||||| ||||| || ||||| || |||||||| ||||||
Db    301 AATTTCACGGTGCCCACAGTTTGGGGTAGCACGGTCCAAGTATCGTGGTTGGCACTCATC 360

Qy    623 CAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCG 682
          || ||||| || |||||||| ||||| || |||||||||||||||  |||||| ||||
Db    361 CATCGCGGTGGCGGCTGCACGTTCGCGGAAAAGATCCATCTGGCTTCCGAGAGAGGGGCT 420

Qy    683 TCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCAC 742
          |||||||| ||||| |||||||||||| ||||| |||||||| |||||||| |||||||||
Db    421 TCTGGAGCGGTCATTTTTAACTTCCCTGGGACGCGCAATGAAGTCATCCCTATGTCTCAC 480

Qy    743 CCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTCTG 802
          |||||||| |  ||||| |||||||| ||||| |||||||||| || ||||||||||||
Db    481 CCGGGTGCGGGGGGACATAGTTGCAATTATGATTGGCAATCTGAAGGGAACAAAAATTCTG 540

Qy    803 CAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGC 862
          || |||||||| |||||||| ||||| |||||||||| |||||||| |||||||| |||
Db    541 CATTCTATTCACAGAGGCATGCAAGTCACAATGGTCATTGAAGTAGGCAAAAAACAGGGC 600

Qy    863 CCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCG 922
          ||||| |||||||| |||||||| || ||||| |||||||| ||||| ||||||||||||
Db    601 CCTTGAGTGAATCATTATTCAATCTTCTTCGTGTCTGTGTCATTTTTCATTATTACGGCC 660

Qy    923 GCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAA 982
          ||||| |||||||||||| |||||||||||||||||| ||  ||||||||||||||||||
Db    661 GCAACCGTGGGCTATTTAATCTTTTATTCTGCTCGCAGATTACGGAATGCAAGAGCTCAA 720

Qy    983 AGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTA 1042
          || |||||||| ||||| || |||||||| |||||||| |||||||| | ||||||  ||
Db    721 AGTAGGAAGCACAGGCAGTTGAAGGCAGAAGCTAAAAATGCTATTGGCAAGCTTCAGCTG 780

Qy   1043 CGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGC 1102
          |||||| ||||||| |||||||| |||||||| |||||||||||||| ||||||||||||
Db    781 CGCACATTGAAACATGGAGACAACGAAATTGGGCCTGATGGAGATAGCTGTGCTGTGTGC 840

Qy   1103 ATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCAT 1162
          |||||  ||||||| |||||||||||| || |||||| |||| || || ||||| ||||||
Db    841 ATTGAGCTGTATAAGCCAAATGATTTTGTGCGCATCCTAACCTGGAATCATATATTCCAT 900

Qy   1163 AAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGAC 1222
          || |||||||| ||||| ||||| ||| |||||| |||||||||||| ||||| || |||

Db    901 AATACATGTGTCGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATCTGCAAGTGGGAC 960

Qy   1223 ATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTC 1282
          || || |||||| | |||||||| |||||||| ||||||||||||||||||||||||||||
Db    961 ATTCTAAAAGCTCTTGGAATTGACGTGGATGTGGAAGATGGATCAGTGTCTTTACAAGTC 1020

Qy   1283 CCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAG 1342
          ||||| || |||||| |||||||| |||||| |||||||||||| | ||| || |||
Db   1021 CCTGTTTCGAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACAGTCGAAGTGAG 1080

Qy   1343 ACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACAC 1402
          || |||||||| |||||||| |||||||| ||| ||||||| || || |||||||| ||
Db   1081 ACTGCATCATCCGGATATGCGTCAGTACAAGGAGCAGATGAGCCACCCCTGGAGGAGCAT 1140

Qy   1403 GTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTG 1462
          | ||| ||| ||||||||| ||| |||||||| |||||||||||||||||||||| ||
Db   1141 GCGCAATCAGCAAATGAAAATCTCCAGCTGGTGAACCATGAAGCAAATTCTGTGGCCGTC 1200

Qy   1463 GATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAG 1522
          |||||| | || ||||| |||||||| |||||||| |||| || |||||| | ||||||
Db   1201 GATGTTGTGCCCCATGTAGACAACCCTACCTTTGACGAAGATGAGACTCCTGAACAAGAG 1260

Qy   1523 ACTGCTGTTCGAGAAATTAAATCTTAA 1549
          |||||||||| || ||||| ||||||
Db   1261 GCTGCTGTTCGCGAGATTAAGTCTTAA 1287
```

Search completed: April 26, 2003, 01:03:20
Job time : 299 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 02:27:30 ; Search time 145 Seconds
(without alignments)
13312.692 Million cell updates/sec

Title: US-09-854-300-7
Perfect score: 1774
Sequence: 1 agctggagctccaccgcggt.....ataggcaagtttcctctcag 1774

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1178.8	66.4	2306	10	US-09-764-864-63	Sequence 63, Appl
	2	1157.6	65.3	1250	10	US-09-764-864-519	Sequence 519, App
	3	372.6	21.0	643	9	US-10-046-935-1815	Sequence 1815, Ap
	4	372.6	21.0	643	9	US-09-878-178-1815	Sequence 1815, Ap
	5	372.6	21.0	643	9	US-10-146-502-1815	Sequence 1815, Ap
c	6	310	17.5	435	10	US-09-920-300A-1176	Sequence 1176, Ap
c	7	310	17.5	435	12	US-10-033-528-1176	Sequence 1176, Ap
	8	274.6	15.5	387	10	US-09-983-965-5551	Sequence 5551, Ap
	9	224.6	12.7	616	9	US-10-060-036-25	Sequence 25, Appl
	10	214.4	12.1	374	10	US-09-983-965-4662	Sequence 4662, Ap
	11	209	11.8	2063	10	US-09-935-390A-1	Sequence 1, Appli
	12	201.2	11.3	392	10	US-09-983-965-5467	Sequence 5467, Ap
	13	187.4	10.6	340	10	US-09-783-590-3523	Sequence 3523, Ap
	14	180.8	10.2	3615	9	US-09-822-846-596	Sequence 596, App
	15	149.2	8.4	1322	10	US-09-935-390A-16	Sequence 16, Appl
	16	132.6	7.5	619	10	US-09-764-864-485	Sequence 485, App
	17	132	7.4	508	9	US-09-918-995-21070	Sequence 21070, A
	18	117.2	6.6	463	10	US-09-864-761-11364	Sequence 11364, A
c	19	112.8	6.4	826	10	US-09-822-849A-90	Sequence 90, Appl
	20	112.4	6.3	435	10	US-09-983-965-5079	Sequence 5079, Ap
	21	105.6	6.0	382	10	US-09-815-343-1328	Sequence 1328, Ap
	22	105.2	5.9	253	10	US-09-864-761-27944	Sequence 27944, A
	23	90.4	5.1	2330	9	US-09-822-846-240	Sequence 240, App
	24	86.4	4.9	922	10	US-09-800-729-23	Sequence 23, Appl
	25	78.8	4.4	570	10	US-09-734-017A-7	Sequence 7, Appli
	26	78.6	4.4	439	9	US-09-925-299-484	Sequence 484, App
	27	78.6	4.4	439	10	US-09-925-299-484	Sequence 484, App
	28	78	4.4	726	9	US-09-986-480-36	Sequence 36, Appl
c	29	77.4	4.4	524	9	US-09-774-639-79	Sequence 79, Appl
c	30	77.4	4.4	524	9	US-09-969-730-81	Sequence 81, Appl
c	31	77.4	4.4	2244	9	US-10-125-540-222	Sequence 222, App
c	32	77.4	4.4	2244	10	US-09-764-870-222	Sequence 222, App
	33	77	4.3	598	9	US-09-764-868-344	Sequence 344, App
c	34	76.8	4.3	1355	9	US-10-125-540-182	Sequence 182, App
c	35	76.8	4.3	1355	10	US-09-764-870-182	Sequence 182, App
c	36	76.8	4.3	1355	10	US-09-764-853-404	Sequence 404, App
	37	76.2	4.3	939	9	US-09-764-868-466	Sequence 466, App
	38	76.2	4.3	940	9	US-09-764-868-40	Sequence 40, Appl
	39	76	4.3	1772	10	US-09-764-853-148	Sequence 148, App
	40	76	4.3	3789	9	US-09-764-868-16	Sequence 16, Appl
	41	75.6	4.3	496	10	US-09-925-301-686	Sequence 686, App
	42	75.6	4.3	1662	9	US-09-764-868-231	Sequence 231, App
	43	75.6	4.3	1906	10	US-09-739-254-32	Sequence 32, Appl
	44	75.6	4.3	1906	10	US-09-904-615-32	Sequence 32, Appl
	45	75.4	4.3	687	9	US-09-774-639-103	Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-764-864-63
; Sequence 63, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-63

Query Match 66.4%; Score 1178.8; DB 10; Length 2306;
Best Local Similarity 99.0%; Pred. No. 4.1e-311;
Matches 1217; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

```
Qy     547 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 71

Qy     607 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 131

Qy     667 TTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 726
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 TTATGAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 191

Qy     727 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     192 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 251

Qy     787 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     252 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 311
```

```
Qy    847 AGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAA-TTTTTTTCGTTTCTGTGTCCT 905
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    312 AGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTTCGTTTCTGTGTCCT 371

Qy    906 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 431

Qy    966 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 491

Qy   1026 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAG 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAG 551

Qy   1086 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 ATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGT 611

Qy   1146 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCC 1205
          |||||||||||||||||||||||||||||||||||||| |||||| ||||||||
Db    612 GCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGCC 671

Qy   1206 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 CCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGAT 731

Qy   1266 CAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAG 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732 CAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAG 791

Qy   1326 AGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAAC 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    792 AGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAAC 851

Qy   1386 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    852 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 911

Qy   1446 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACG 1505
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    912 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGACG 971

Qy   1506 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    972 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1031

Qy   1566 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAAT 1625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1032 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAAT 1091

Qy   1626 TGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCT 1685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 TGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCT 1151

Qy   1686 ATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATA 1745
          |||||||||   |  | |||||||||||||||||||||||||| ||||||||||||
Db   1152 ATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATA 1208

Qy   1746 ATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          |||||||||||||||||||||||||||||
Db   1209 ATATTTTTCATAGGCAAGTTTCCTCTCAG 1237


RESULT 2
US-09-764-864-519
; Sequence 519, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 519
;   LENGTH: 1250
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (540)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1242)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-519

  Query Match             65.3%;  Score 1157.6;  DB 10;  Length 1250;
  Best Local Similarity   98.8%;  Pred. No. 1.7e-305;
  Matches 1208;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    4;

Qy    547 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     18 CTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTC 77

Qy    607 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     78 TTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGC 137

Qy    667 TTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 726
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    138 TTATGAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGT 197

Qy    727 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    198 CATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAA 257

Qy    787 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    258 AGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGT 317

Qy    847 AGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAA-TTTTTTTCGTTTCTGTGTCCT 905
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    318 AGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTTCGTTTCTGTGTCCT 377

Qy    906 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 TTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTAC 437

Qy    966 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 GGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTA 497

Qy   1026 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGG-AAATTGGCCCTGATGGA 1084
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    498 TTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGNAAATTGGCCCTGATGGA 557

Qy   1085 GATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACG 1144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 GATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACG 617

Qy   1145 TGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGC 1204
          |||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db    618 TGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGACTTGC 677

Qy   1205 CCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGA 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 CCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGA 737

Qy   1265 TCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAA 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     738 TCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAA 797

Qy    1325 GAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAA 1384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 GAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAA 857

Qy    1385 CCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAA 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 CCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAA 917

Qy    1445 GCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGAC 1504
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     918 GCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAGAAGAC 977

Qy    1505 GAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAG 1564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 GAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAG 1037

Qy    1565 AAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAA 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 AAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAA 1097

Qy    1625 TTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGC 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 TTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGC 1157

Qy    1685 TATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTAT 1744
           ||||||||||    |  | ||||||||||||||||||||||||||| ||||||||||||
Db    1158 TATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTAT 1214

Qy    1745 AATATTTTTCATAGGCAAGTTTC 1767
           ||||||||||||||||||  || |
Db    1215 AATATTTTTCATAGGCAATTTCC 1237
```

RESULT 3
US-10-046-935-1815
; Sequence 1815, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1815
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 584
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1815

Query Match 21.0%; Score 372.6; DB 9; Length 643;
Best Local Similarity 98.0%; Pred. No. 2.1e-91;
Matches 398; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

```
Qy    1369 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 1428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 60

Qy    1429 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCC 1488
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      61 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCC 120

Qy    1489 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 1548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 180

Qy    1549 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 1608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 240

Qy    1609 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 1668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 300

Qy    1669 ATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTT 1728
           |||||||||||||||||||||||||    |  | ||||||||||||||||||||||||||
Db     301 ATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTT 358

Qy    1729 CCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
           ||| ||||||||||||||||||||||||||||||||||||||||||
Db     359 CCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 403
```

RESULT 4
US-09-878-178-1815
; Sequence 1815, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1815
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(643)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1815

Query Match 21.0%; Score 372.6; DB 9; Length 643;
Best Local Similarity 98.0%; Pred. No. 2.1e-91;
Matches 398; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

```
Qy    1369 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 1428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 60

Qy    1429 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCC 1488
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      61 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCC 120

Qy    1489 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 1548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 180

Qy    1549 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 1608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 240

Qy    1609 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 1668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 300

Qy    1669 ATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTT 1728
           |||||||||||||||||||||||||    |  | ||||||||||||||||||||||||||
```

```
Db    301 ATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTT 358

Qy   1729 CCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          ||| |||||||||||||||||||||||||||||||||||||||||
Db    359 CCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 403


RESULT 5
US-10-146-502-1815
; Sequence 1815, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.527C2
;  CURRENT APPLICATION NUMBER: US/10/146,502
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 2241
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1815
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 584
;   OTHER INFORMATION: n = A,T,C or G
US-10-146-502-1815

  Query Match             21.0%;  Score 372.6;  DB 9;  Length 643;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-91;
  Matches  398;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    2;

Qy   1369 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 1428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACA 60

Qy   1429 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCC 1488
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     61 GCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCC 120

Qy   1489 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 1548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTA 180

Qy   1549 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 1608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCT 240

Qy   1609 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 1668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAG 300

Qy   1669 ATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTT 1728
          |||||||||||||||||||||||||    |  | ||||||||||||||||||||||||||||
Db    301 ATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTT 358

Qy   1729 CCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          ||| |||||||||||||||||||||||||||||||||||||||||
Db    359 CCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 403


RESULT 6
US-09-920-300A-1176/c
; Sequence 1176, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547
;  CURRENT APPLICATION NUMBER: US/09/920,300A
;  CURRENT FILING DATE:  2001-07-31
;  NUMBER OF SEQ ID NOS: 1789
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1176
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-920-300A-1176

  Query Match             17.5%;  Score 310;  DB 10;  Length 435;
  Best Local Similarity   97.4%;  Pred. No. 2e-74;
  Matches  336;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    2;

Qy   1430 CTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCA 1489
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    435 CTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCA 376

Qy   1490 ACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAA 1549
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    375 ACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGTTGTTCGAGAAATTAAATCTTAA 316

Qy   1550 AATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTA 1609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 AATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTA 256

Qy   1610 GTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGA 1669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGA 196

Qy   1670 TGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTC 1729
          |||||||||||||||||||||||||    |  | ||||||||||||||||||||||||||||
Db    195 TGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTC 138

Qy   1730 CACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          || |||||||||||||||||||||||||||||||||||||||||||
Db    137 CA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 94


RESULT 7
US-10-033-528-1176/c
; Sequence 1176, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547C1
;  CURRENT APPLICATION NUMBER: US/10/033,528
;  CURRENT FILING DATE:  2001-12-26
;  NUMBER OF SEQ ID NOS: 1896
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1176
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-033-528-1176

  Query Match             17.5%;  Score 310;  DB 12;  Length 435;
  Best Local Similarity   97.4%;  Pred. No. 2e-74;
  Matches  336;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    2;

Qy   1430 CTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCA 1489
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
```

```
Db    435 CTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCA 376

Qy   1490 ACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAA 1549
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    375 ACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGTTGTTCGAGAAATTAAATCTTAA 316

Qy   1550 AATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTA 1609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 AATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTA 256

Qy   1610 GTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGA 1669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGA 196

Qy   1670 TGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTC 1729
          |||||||||||||||||||||||||    | |||||||||||||||||||||||||||||
Db    195 TGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTC 138

Qy   1730 CACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          || ||||||||||||||||||||||||||||||||||||||||||
Db    137 CA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 94
```

RESULT 8
US-09-983-965-5551
; Sequence 5551, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5551
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (347)
; OTHER INFORMATION:
; OTHER INFORMATION: Clone ID: 50-LIB34-010-Q1-E2-E6
US-09-983-965-5551

Query Match 15.5%; Score 274.6; DB 10; Length 387;
Best Local Similarity 86.8%; Pred. No. 8.5e-65;
Matches 336; Conservative 0; Mismatches 46; Indels 5; Gaps 3;

```
Qy   1386 CGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAG 1445
          | ||| |||||||||| |  || ||| |||||||| ||||||||||||||||||||||
Db       1 CACCTTTGGAGGAACATGCACACTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAG 60

Qy   1446 CAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACG 1505
          |||||||  |||||||||||||| ||||||| || |||||||||||||||||||||||| |
Db      61 CAAATTCCATGGCAGTGGATGTTGTTCCTCACGTTGACAACCCAACCTTTGAAGAAGATG 120

Qy   1506 AAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGA 1565
          |||  |||| |||| |||||  |||||||||||||||||||| |||||||||| | | ||||
Db    121 AAAGTCCTGATCAGGAGACCACTGTTCGAGAAATTAAATCATAAAATCTGTATCAGTAGA 180

Qy   1566 AAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTT-CTATTAATAAA 1624
          |||| |||||| |||||||||||||||||||||||||||||||||| ||| |||  ||
Db    181 AAACCTGAACCTTTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTACTACTAAGGAA 240

Qy   1625 TTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGC 1684
          |||||||| |||||||||||||| |||||||||||||||| | ||||||||||||||||
Db    241 CTGGATAAACTTAATAAAATAAGAATGATACTGAAAGTGCTGATATGACTAATATTATGC 300

Qy   1685 TATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTAT 1744
          ||||||||||    |  | |||||| | ||||| ||||| |||| |  || || ||||||
Db    301 TATAGTTAAA--TGACTTAAAATACTAAACCTATTAACCTTTT--CCACANACACATTAT 356

Qy   1745 AATATTTTTCATAGGCAAGTTTCCTCT 1771
          ||| |||||||||||||||||||||||
Db    357 AATGTTTTTCATAGGCAAGTTTCCTCT 383
```

RESULT 9
US-10-060-036-25
; Sequence 25, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 594
; OTHER INFORMATION: n = A,T,C or G
US-10-060-036-25

Query Match 12.7%; Score 224.6; DB 9; Length 616;
Best Local Similarity 97.3%; Pred. No. 5.2e-51;
Matches 250; Conservative 0; Mismatches 4; Indels 3; Gaps 2;

```
Qy   1518 AAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCA 1577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCA 60

Qy   1578 TTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATTTA 1637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATTTA 120

Qy   1638 ATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAAAATG 1697
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAAA--T 178

Qy   1698 GCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATATTTTTCATA 1757
          |  | |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    179 GGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTCATTATAATATTTTTCATA 237

Qy   1758 GGCAAGTTTCCTCTCAG 1774
          |||||||||||||||||
Db    238 GGCAAGTTTCCTCTCAG 254
```

RESULT 10
US-09-983-965-4662
; Sequence 4662, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing

```
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION:  NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4662
;   LENGTH: 374
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 14-LIB34-009-Q1-E1-D5
US-09-983-965-4662

  Query Match             12.1%;  Score 214.4;  DB 10;  Length 374;
  Best Local Similarity   78.2%;  Pred. No. 2.3e-48;
  Matches  283;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    2;

Qy      75 CGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACGT 134
Db      13 CAAAGGGCTTCCTCTGCGGCAACATCTCCGGCGACGCTACGTGCCCGCCGGCTCTGACCC 72

Qy     135 AGCTCGCAGCTCC--CCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAA 192
Db      73 AGCCCGCCGCGCGCGCCAGTCTCGTCCCGTCCTCTCACTACCTGGCGCGCACCTGGCCCA 132

Qy     193 GACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAG 252
Db     133 GGCCGGGGTCCCGCCGAGAGTCGGGAAGGCGCGTGCCAGGGGCCCTAGGGAACCTCAGAG 192

Qy     253 CGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTT 312
Db     193 CGCGCATATCATGGGGCAGCTGCCAGGGGCCGGGGTCTTCTGCCGCGGTGGTTGTGGCTT 252

Qy     313 TTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCG 372
Db     253 TTCCCGATTGCTGGCCTGGGGCTTCCTGCTGGTTCTGAGTCCGCAAACCTCCGGCTCCAG 312

Qy     373 GGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGT-GTCCTGGCGGGTTCCGCACACGG 431
Db     313 TGGAGCCGGAGCCGTGTGGGACGCGTAGCTCAACGTGGTCCTGGCGGGTTCCGCACACCG 372

Qy     432 GA 433
Db     373 GA 374

RESULT 11
US-09-935-390A-1
; Sequence 1, Application US/09935390A
; Patent No. US20020076761A1
;  GENERAL INFORMATION:
;        APPLICANT: Escobedo, Jaime
;                   Quianjin, Hu
;                   Garcia, Pablo
;                   Williams, Lewis T.
;                   Kothakota, Srinivas
;        TITLE OF INVENTION: Secreted Human Proteins
;        NUMBER OF SEQUENCES: 38
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Chiron Corporation
;             STREET: 4560 Horton Street
;             CITY: Emeryville
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94608-2916
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/935,390A
;             FILING DATE: 22-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/988,671
;             FILING DATE: 1997-12-11
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jane E. R. Potter
;             REGISTRATION NUMBER: 33,332
;             REFERENCE/DOCKET NUMBER: 1369.002
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (510) 923-2718
;             TELEFAX: (510) 655-3542
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2063 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-935-390A-1

  Query Match             11.8%;  Score 209;  DB 10;  Length 2063;
  Best Local Similarity   54.6%;  Pred. No. 2.1e-46;
  Matches  507;  Conservative    0;  Mismatches  410;  Indels   12;  Gaps    4;

Qy     339 TGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGGGGGGCTGAAGCAGTGTGGACCGCGT 398
Db     180 TGCTCGCCCTGGCCCTGTGCGTGCCCGGGGCCCGGGGCCGGGCTCTCGAGTGGTTCTCGG 239

Qy     399 ACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGAGTGAACCGTACGGTGTGGGAGCTGA 458
Db     240 CCGT---GGTAAACATCGAGTACGTGGACCCGCAGACCAACCTGACGGTGTGGAGCGTCT 296

Qy     459 GCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTGTGGCTGGGGTCCTGGTAC 518
Db     297 CGGAGAGTGGCCGCTTCGGCGACAGCTCGCCCAAGGAGGGCGCGCATGGCCTGGTGGGCG 356

Qy     519 CGCCCGACGGCCCGGGGCGCTTAACGCCTGTAACCCGCACACGAATTTCACGGTGCCCA 578
Db     357 TCCCGTGGGCGCCCGGCGGAGACCTCGAGGGCTGCGCGCCCGACACGCGCTTCTTCGTGC 416

Qy     579 CGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCATCCAACGCGGCGGGGGCT 638
Db     417 CCGAGCCCGGCGGCCGAGGGGCCGCGCCCTGGGTCGCCCTGGTGGCTC---GTGGGGGCT 473

Qy     639 GCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGGCGTCTGGAGCCGTCATCT 698
Db     474 GCACCTTCAAGGACAAGGTGCTGGTGGCGGCGCGGAGGAACGCCTCGGCCGTCGTCCTCT 533

Qy     699 TTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACA 758
Db     534 ACAATGAGGAGCGCTACGGGAACATCACCTTGCCCATGTCTCACGCGGGAACAGGAAATA 593

Qy     759 TTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAAATTCTGCAATCTATTCAAAGAG 818
Db     594 TAGTGGTCATTATGATTAGCTATCCAAAAGGAAGAGAAATTTTGGAGCTGGTGCAAAAAG 653

Qy     819 GCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATGGCCCTTGGGTGAATCAC- 877
Db     654 GAATTCCAGTAACGATGACCATAGGGGTTGGCACCCGGCATGTACAGGAGTTCATCAGCG 713

Qy     878 --TATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCT 935
Db     714 GTCAGTCTGTGGTGTTTGTGGCCATTGCCTTCATCACCATGATGATTATCTCGTTAGCCT 773
```

```
Qy     936 ATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGA 995
           | || ||||| | |    |  | | |   || |  |  |||| |   | | |||| 
Db     774 GGCTAATATTTTACTATATACAGCGTTTCCTATATACTGGCTCTCAGATTGGAAGTCAGA 833

Qy     996 GGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAAC 1055
           | ||   |||   |||  |||| |||| ||||||   ||||| ||| |  ||  | || |
Db     834 GCCATAGAAA---AGAAACTAAGAAAGTTATTGGCCAGCTTCTACTTCATACTGTAAAGC 890

Qy    1056 AAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATA 1115
           | ||||| |||| |||||     |||||  || | |||||| ||||| ||||||   |  |
Db     891 ATGGAGAAAAGGGAATTGATGTTGATGCTGAAAATTGTGCAGTGTGTATTGAAAATTTCA 950

Qy    1116 AACCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTG 1175
           ||  ||| ||| |  |  | ||  |  | ||||| |||||||| ||||  | |||  |||
Db     951 AAGTAAAGGATATTATTAGAATTGTGCCATGCAAGCATATTTTTCATAGAATATGCATTG 1010

Qy    1176 ACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTT 1235
           ||||||||||  ||  | |||   || || || ||||| |||  |||  |  |||||||
Db    1011 ACCCATGGCTTTTGGATCACCGAACATGTCCAATGTGTAAACTTGATGTCATCAAAGCCC 1070

Qy    1236 TGGGAATTGAGGTGGATGTTGAAGATGGA 1264
           | |||  |  ||  ||   ||  |||| |
Db    1071 TAGGATATTGGGGAGAGCCTGGGGATGTA 1099


RESULT 12
US-09-983-965-5467
; Sequence 5467, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5467
;   LENGTH: 392
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 46-LIB34-036-Q1-E1-D10
US-09-983-965-5467

  Query Match           11.3%;  Score 201.2;  DB 10;  Length 392;
  Best Local Similarity   80.3%;  Pred. No. 9.3e-45;
  Matches  285;  Conservative    0;  Mismatches   64;  Indels    6;  Gaps    4;

Qy    1418 GAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCAT 1477
           |||| |||||||||||||||||||||||||||||   ||||||||||||| |||||||
Db       1 GAAAATCTACAGCTGGTAAACCATGAAGCAAATTCCATGGCAGTGGATGTTGTTCCTCAC 60

Qy    1478 GTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAA 1537
           || |||||||||||||||||||||| |||| |||| |||| || ||   || || ||||
Db      61 GTTGACAACCCAACCTTTGAAGAAGATGAAAGTCCTGATCAGGAAAC-AATGGTCCAGAA 119

Qy    1538 ATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCC 1597
           |||||||| ||||||| || | ||| |||||  |||||| ||| |||||||| ||||| |
Db     120 ATTAAATCATAAAATCCGTATCAATTGAAAAACTGAACCTTTAATAATAACAAAACTGGC 179

Qy    1598 AATC-AGGGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACT 1656
           ||||  |||||| | || | |  ||   | ||| ||||  |   | ||||  ||  || |||
Db     180 AATCAAGGGCCAATTTACAACAAAGGGAATGGGTAAACCTTTAATAATATGAAAGAGACT 239

Qy    1657 GAAAGTGCTCAGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCT 1716
           |||| |||| | ||||||||||||||||||||||||||   |  | |||||| | |||||
Db     240 GAAAATGCTGATATGACTAATATTATGCTATAGTTAAA--TGACTTAAAATACTAAACCT 297

Qy    1717 GTTAACTTTTTTCCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCT 1771
           |||||  ||||  |  ||||| ||||||||  ||||||||||||||| ||||||||
Db     298 ATTAACCTTTT--CCACAAACACATTATAAAGTTTTTCATAGGCAAATTTCCTCT 350


RESULT 13
US-09-783-590-3523
; Sequence 3523, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3523
;   LENGTH: 340
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (205)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (220)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (221)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (244)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (265)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (294)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (318)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (326)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (332)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (333)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3523

  Query Match           10.6%;  Score 187.4;  DB 10;  Length 340;
  Best Local Similarity   97.0%;  Pred. No. 5e-41;
  Matches  191;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    1203 GCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATG 1262
           | | |  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGCACGAGGGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATG 60
```

```
Qy   1263 GATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATG 1322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATG 120

Qy   1323 AAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATG 1382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATG 180

Qy   1383 AACCGCCTCTGGAGGAA 1399
          |||||||||||||||| |
Db    181 AACCGCCTCTGGAGGTA 197
```

RESULT 14
US-09-822-846-596
; Sequence 596, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 596
; LENGTH: 3615
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-596

Query Match 10.2%; Score 180.8; DB 9; Length 3615;
Best Local Similarity 57.2%; Pred. No. 1.5e-38;
Matches 348; Conservative 0; Mismatches 257; Indels 3; Gaps 1;

```
Qy    653 AAGATCCATCTGGCTTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGG 712
          |||||||     || |     |        || || |   |  |||||||| ||| |    |
Db      1 AAGATCCGGAACGCGTTCCTGCAGAACGCCTCAGCCGTGGTCATCTTCAACGTGGGCTCC 60

Qy    713 ACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATG 772
          | |  ||| |||  |||| ||||| | ||| |||||| || |||||| || || || |||
Db     61 AACACCAACGAGACCATCACCATGCCCCACGCGGGTGTAGAAGACATCGTGGCCATAATG 120

Qy    773 ATCGGCAATCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACA 832
          ||     | |  ||||| |   | ||  |          |  |||||  |||    ||||||
Db    121 ATTCCTGAGCCAAAAGGGAAGGAGATAGTAAGCCTGCTGGAAAGAAACATCACCGTGACA 180

Qy    833 ATGGTCATAGAAGTAGGGA---AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTT 889
          |||  |||     | || |     ||        |  ||||  | |  |||  || |
Db    181 ATGTACATCACCATCGGAACCCGGAACTTGCAGAAATATGTGAGCCGCACTTCGGTTGTG 240

Qy    890 TTCGTTTCTGTGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTAT 949
          || || ||  | |||||  || |  | | |       |  | |  |   |  ||||||||
```

```
Db    241 TTTGTCTCCATCTCCTTCATTGTCCTGATGATCATTTCCCTCGCATGGCTCGTCTTTTAT 300

Qy    950 TCTGCTCGAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCA 1009
          |      | ||| | ||  ||||||  ||        |||||  ||| | | | |   ||
Db    301 TACATCCAGAGGTTTCGATATGCAAATGCCAGGGATAGGAACCAGCGCCGACTGGGGGAT 360

Qy   1010 GATGCTAAAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAA 1069
          |  || || ||||| ||  | |  || ||  |  | ||  | ||  | || |||||||||
Db    361 GCAGCAAAGAAAGCCATCAGCAAACTCCAGATCAGGACCATCAAGAAGGGTGACAAGGAA 420

Qy   1070 ATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTG 1129
          |  |   |||||   || |  ||||| || || ||||||  ||| || || |||||   |
Db    421 ACAGAGTCTGATTTTGACAACTGTGCAGTTTGTATTGAAGGGTACAAGCCCAATGACGTT 480

Qy   1130 GTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTA 1189
          || || ||| |  | |||   ||| |||||| ||| | |||||||||| |||||  ||
Db    481 GTCCGGATCCTGCCCTGCCGGCATCTTTTCCACAAGTCCTGTGTTGACCCCTGGCTTCTA 540

Qy   1190 AAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTG 1249
          | ||    || || ||||||||||      |||| || ||||  | || ||   |
Db    541 GACCATCGTACCTGTCCCATGTGCAAGATGAACATTCTTAAAGCCCTAGGGATCCCGCCC 600

Qy   1250 GATGTTGA 1257
          |||  ||
Db    601 AATGCCGA 608
```

RESULT 15
US-09-935-390A-16
; Sequence 16, Application US/09935390A
; Patent No. US20020076761A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; Quianjin, Hu
; Garcia, Pablo
; Williams, Lewis T.
; Kothakota, Srinivas
; TITLE OF INVENTION: Secreted Human Proteins
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/935,390A
; FILING DATE: 22-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/988,671
; FILING DATE: 1997-12-11
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. R. Potter
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1369.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1322 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:

US-09-935-390A-16

Query Match 8.4%; Score 149.2; DB 10; Length 1322;
Best Local Similarity 56.1%; Pred. No. 3.3e-30;
Matches 302; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

```
Qy    720 ATGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCA 779
          | |||     |  ||||| |||| || ||   | | ||| ||| ||||  |||||||
Db    259 AGGAGCCAGTTACCATGACTCATCCAGGCACTGGAGATATTATTGCTGTCATGATAACAG 318

Qy    780 ATCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCA 839
          |  |||  || |   | ||| ||   | | |  | | |  |||    ||   ||||   |
Db    319 AATTGAGGGGTAAGGATATTTTGAGTTATCTGGAGAAAAACATCTCTGTACAAATGACAA 378

Qy    840 TAGAAGTAGGGAAAAAACATGGCGCTTGGG---TGAATCACTATTCAATTTTTTTCGTTT 896
          |||  || || |   |    ||   |  | |  |    ||  |  | ||||| |
Db    379 TAGCTGTTGGAACTCGAATGCCACCGAAGAACTTCAGCCGTGGCTCTCTAGTCTTCGTGT 438

Qy    897 CTGTGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTC 956
          |  | ||||||  || || | | |        ||   |  |   | || || || |    ||
Db    439 CAATATCCTTTATTGTTTTGATGATTATTTCTTCAGCATGGCTCATATTCTACTTCATTC 498

Qy    957 GAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTA 1016
           || | |  || |  |||  || |    ||||| |||  | |   |    |  |  || |
Db    499 AAAAGATCAGGTACACAAATGCACGCGACAGGAACCAGCGTCGTCTCGGAGATGCAGCCA 558

Qy   1017 AAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCC 1076
          | ||||| ||  | |   |    |   | ||| | ||  | || |||||||||| || ||
Db    559 AGAAAGCCATCAGTAAATTGACAACCAGGACAGTAAAGAAGGGTGACAAGGAAACTGACC 618

Qy   1077 CTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCA 1136
          | ||    |||  |||||| || ||||| ||    ||||| |  |||||| | || || |
Db    619 CAGACTTTGATCATTGTGCAGTCTGCATAGAGAGCTATAAGCAGAATGATGTCGTCCGAA 678

Qy   1137 TCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACA 1196
          |  |  | ||||| ||| |||||||| ||  | || || || || |||||     ||||
Db    679 TTCTCCCCTGCAAGCATGTTTTCCACAAATCCTGCGTGGATCCCTGGCTTAGTGAACATT 738

Qy   1197 AGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGT 1254
           || || || |||||||||  | | ||| | || ||  |||||||| |  | || |
Db    739 GTACCTGTCCTATGTGCAAACTTAATATATTGAAGGCCCTGGGAATTGTGCCGAATTT 796
```

Search completed: April 26, 2003, 03:26:46
Job time : 164 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 00:58:15 ; Search time 57 Seconds
(without alignments)
9544.642 Million cell updates/sec

Title: US-09-854-300-7
Perfect score: 1774
Sequence: 1 agctggagctccaccgcggt.....ataggcaagtttcctctcag 1774

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	147.6	8.3	1253	2	US-08-786-606-6	Sequence 6, Appli
2	77	4.3	912	2	US-09-090-567-3	Sequence 3, Appli
3	76.4	4.3	2605	2	US-08-680-395-4	Sequence 4, Appli
4	73.8	4.2	1378	1	US-08-075-533-20	Sequence 20, Appl
5	73.8	4.2	1378	2	US-08-948-176-20	Sequence 20, Appl
6	73.8	4.2	1378	5	PCT-US91-09160-20	Sequence 20, Appl
7	73.8	4.2	1556	4	US-09-043-937A-3	Sequence 3, Appli
8	73.8	4.2	1568	4	US-09-043-937A-1	Sequence 1, Appli
9	73.2	4.1	1543	4	US-09-227-357-26	Sequence 26, Appl
10	73	4.1	545	4	US-09-227-357-125	Sequence 125, App
11	73	4.1	1585	4	US-09-183-861-54	Sequence 54, Appl
12	73	4.1	1585	4	US-09-022-765-54	Sequence 54, Appl
13	73	4.1	2775	1	US-08-730-771-1	Sequence 1, Appli
14	73	4.1	2775	4	US-09-060-208-1	Sequence 1, Appli
15	73	4.1	3089	1	US-08-472-934-5	Sequence 5, Appli
16	73	4.1	3089	2	US-08-323-460A-5	Sequence 5, Appli
17	73	4.1	3089	2	US-08-461-146C-5	Sequence 5, Appli
18	73	4.1	3089	3	US-08-461-145C-5	Sequence 5, Appli
19	73	4.1	3089	4	US-08-628-829-9	Sequence 9, Appli
20	73	4.1	3328	4	US-08-960-048-1	Sequence 1, Appli
21	71.4	4.0	2045	4	US-08-795-088A-1	Sequence 1, Appli
22	71	4.0	625	4	US-09-288-143-17	Sequence 17, Appl
23	70.4	4.0	2085	2	US-08-668-128B-7	Sequence 7, Appli
24	70.4	4.0	2085	2	US-08-905-445-7	Sequence 7, Appli
25	69.4	3.9	633	4	US-08-818-112-6	Sequence 6, Appli
26	69.4	3.9	633	4	US-08-818-111-6	Sequence 6, Appli
27	69.4	3.9	633	4	US-09-056-556-6	Sequence 6, Appli
28	69.4	3.9	633	4	US-09-072-596-6	Sequence 6, Appli
c 29	68.2	3.8	500	4	US-08-818-112-101	Sequence 101, App
c 30	68.2	3.8	500	4	US-08-818-111-96	Sequence 96, Appl
c 31	68.2	3.8	500	4	US-09-056-556-101	Sequence 101, App
c 32	68.2	3.8	500	4	US-09-072-596-96	Sequence 96, Appl
33	67.8	3.8	2418	4	US-09-285-379-1	Sequence 1, Appli
34	66.6	3.8	2327	4	US-09-149-476-107	Sequence 107, App
c 35	66.2	3.7	3023	4	US-09-308-022-4	Sequence 4, Appli
c 36	66	3.7	752	4	US-08-976-259-108	Sequence 108, App
37	66	3.7	849	3	US-08-646-538-14	Sequence 14, Appl
38	66	3.7	849	4	US-09-503-222-14	Sequence 14, Appl
c 39	66	3.7	1924	4	US-09-424-283-5	Sequence 5, Appli
c 40	66	3.7	1958	4	US-09-215-221-9	Sequence 9, Appli
c 41	66	3.7	2961	4	US-08-446-935-6	Sequence 6, Appli
c 42	66	3.7	3699	3	US-08-646-538-6	Sequence 6, Appli
c 43	66	3.7	3699	4	US-09-503-222-6	Sequence 6, Appli
c 44	66	3.7	3792	2	US-08-992-334-1	Sequence 1, Appli
c 45	66	3.7	3792	3	US-08-302-752-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-786-606-6
; Sequence 6, Application US/08786606
; Patent No. 5861495
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Au-Young, Janice
; APPLICANT: Coleman, Roger
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,606
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0173 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1253 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-786-606-6

Query Match 8.3%; Score 147.6; DB 2; Length 1253;
Best Local Similarity 55.9%; Pred. No. 2.6e-31;
Matches 301; Conservative 0; Mismatches 234; Indels 3; Gaps 1;

```
Qy    720 ATGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCA 779
          | |||     |  ||||| |||| || ||  | | ||| ||| ||||  |||||||
Db    242 AGGAGCCAGTTACCATGACTCATCCAGGCACTGGAGATATTATTGCTGTCATGATAACAG 301

Qy    780 ATCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCA 839
          |  |||  || |    | ||| ||    | | |  | | |  |||     ||   ||||   |
Db    302 AATTGAGGGGTAAGGATATTTTGAGTTATCTGGAGAAAAACATCTCTGTACAAATGACAA 361

Qy    840 TAGAAGTAGGGAAAAAACATGGCCCTTGGG---TGAATCACTATTCAATTTTTTTCGTTT 896
          |||  || || |     |         ||   |    | | |       ||  |  | ||||| |
Db    362 TAGCTGTTGGAACTCGAATGCCACCGAAGAACTTCAGCCGTGGCTCTCTAGTCTTCGTGT 421

Qy    897 CTGTGTCCTTTTTTATTATTAGGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTC 956
          |  | |||||| || || | | |          ||   | |   | || || || |    ||
Db    422 CAATATCCTTTATTGTTTTGATGATTATTTCTTCAGCATGGCTCATATTCTACTTCATTC 481

Qy    957 GAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTA 1016
            | |  || |  |||  || |    ||||| ||| | |   |    |  |  || |
Db    482 AGAAGATCAGGTACACAAATGCACGCGACAGGAACCAGCGTCGTCTCGGAGATGCAGCCA 541

Qy   1017 AAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCC 1076
          | ||||| ||  | |   |   |    | ||| | ||  | || ||||||||||| || ||
Db    542 AGAAAGCCATCAGTAAATTGACAACCAGGACAGTAAAGAAGGGTGACAAGGAAACTGACC 601

Qy   1077 CTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCA 1136
          | ||    |||  |||||| || ||||| ||     ||||| |  |||||| | || || |
Db    602 CAGACTTTGATCATTGTGCAGTCTGCATAGAGAGCTATAAGCAGAATGATGTCGTCCGAA 661

Qy   1137 TCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACA 1196
          |  |  | ||||| ||| ||||||| ||  | || || || || |||||      ||||
Db    662 TTCTCCCCTGCAAGCATGTTTTCCACAAATCCTGCGTGGATCCCTGGCTTAGTGAACATT 721

Qy   1197 AGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGT 1254
            || || || |||||||||  | | ||| | || ||  |||||||| |  | || |
Db    722 GTACCTGTCCTATGTGCAAACTTAATATATTGAAGGCCCTGGGAATTGTGCCGAATTT 779
```

RESULT 2
US-09-090-567-3
; Sequence 3, Application US/09090567
; Patent No. 5989549
; GENERAL INFORMATION:
; APPLICANT: Sullivan, Robert
; APPLICANT: Brub, Bruno
; APPLICANT: Lgar, Christine
; APPLICANT: Gaudreault, Christian
; TITLE OF INVENTION: Acrosomal Sperm Protein And
; TITLE OF INVENTION: Uses Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swabey Ogilvy Renault
; STREET: 1600 - 1981 McGill College
; CITY: Monteal
; STATE: QC
; COUNTRY: Canada
; ZIP: H3A 2Y3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/090,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Kevin P
; REGISTRATION NUMBER: 26,674
; REFERENCE/DOCKET NUMBER: 13045-2"US" FC/CC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 514-845-7126
; TELEFAX: 514-288-8389
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
US-09-090-567-3

Query Match 4.3%; Score 77; DB 2; Length 912;
Best Local Similarity 94.1%; Pred. No. 1.1e-11;
Matches 80; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 66

Qy     61 AATTCGGCACGAGCCGAGGAGCTGC 85
          |||||||||||||||||    | ||
Db     67 AATTCGGCACGAGCCGACATGGAGC 91
```

RESULT 3
US-08-680-395-4
; Sequence 4, Application US/08680395
; Patent No. 5892010
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W.
; APPLICANT: Collins, Colin
; APPLICANT: Hwang, Soo-in
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowbel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,395
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 02307O-068900US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: -

```
;       LOCATION:  1..2605
;       OTHER INFORMATION:  /note= "cDNA clone cc43 of 4 kb
;       OTHER INFORMATION:  transcript"
US-08-680-395-4

  Query Match             4.3%;  Score 76.4;  DB 2;  Length 2605;
  Best Local Similarity   80.9%;  Pred. No. 2.8e-11;
  Matches   89;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 95

Qy      61 AATTCGGCACGAGCCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACG 110
           |||||||||||||| | |   |||| |||   | | | | | ||
Db      96 AATTCGGCACGAGCTGGGCTACTACGATGGCGATGAGTTTCGAGTGGCCG 145


RESULT 4
US-08-075-533-20
; Sequence 20, Application US/08075533
; Patent No. 5530186
;  GENERAL INFORMATION:
;    APPLICANT:  Hitz, William D.
;    APPLICANT:  Yadav, Narendra S.
;    TITLE OF INVENTION:  Nucleotide Sequences of Soybean Acyl-ACP
;    TITLE OF INVENTION:  Thioesterase Genes
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. du Pont de Nemours and Company
;      STREET:  1007 Market Street
;      CITY:  Wilmington
;      STATE:  Delaware
;      COUNTRY:  U.S.A
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,533
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/631,264
;      FILING DATE:  20-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Morrissey, Bruce W.
;      REGISTRATION NUMBER:  30,663
;      REFERENCE/DOCKET NUMBER:  CR-8926-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (302) 992-4927
;      TELEFAX:  (302) 892-7949
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Brassica napus
US-08-075-533-20

  Query Match             4.2%;  Score 73.8;  DB 1;  Length 1378;
  Best Local Similarity   97.4%;  Pred. No. 1.1e-10;
  Matches   75;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 111

Qy      61 AATTCGGCACGAGCCGA 77
           |||||||||||||  ||
Db     112 AATTCGGCACGAGAAGA 128


RESULT 5
US-08-948-176-20
; Sequence 20, Application US/08948176
; Patent No. 5945585
;  GENERAL INFORMATION:
;    APPLICANT:  HITZ, WILLIAM D.
;    APPLICANT:  YADAV, NARENDRA S.
;    TITLE OF INVENTION:  ACYL-ACP THIOESTERASES GENES
;    TITLE OF INVENTION:  AND THEIR USE IN ALTERING PLANT
;    TITLE OF INVENTION:  OIL COMPOSITION
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,176
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/631,264
;      FILING DATE:  DECEMBER 20, 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHRISTENBURY, LYNNE M.
;      REGISTRATION NUMBER:  30,971
;      REFERENCE/DOCKET NUMBER:  CR-8926-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-5481
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Brassica napus
US-08-948-176-20

  Query Match             4.2%;  Score 73.8;  DB 2;  Length 1378;
  Best Local Similarity   97.4%;  Pred. No. 1.1e-10;
  Matches   75;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 111

Qy      61 AATTCGGCACGAGCCGA 77
           |||||||||||||  ||
Db     112 AATTCGGCACGAGAAGA 128


RESULT 6
PCT-US91-09160-20
; Sequence 20, Application PC/TUS9109160
;  GENERAL INFORMATION:
;    APPLICANT:  Hitz, William D.
;    APPLICANT:  Yadav, Narendra S.
```

```
;    TITLE OF INVENTION:  Nucleotide Sequences of Soybean Acyl-ACP
;    TITLE OF INVENTION:  Thioesterase Genes
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. du Pont de Nemours and Company
;      STREET:  1007 Market Street
;      CITY:  Wilmington
;      STATE:  Delaware
;      COUNTRY:  U.S.A
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09160
;      FILING DATE:  19911216
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/631,264
;      FILING DATE:  20-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Morrissey, Bruce W.
;      REGISTRATION NUMBER:  30,663
;      REFERENCE/DOCKET NUMBER:  CR-8926-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (302) 992-4927
;      TELEFAX:  (302) 892-7949
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1378 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Brassica napus
PCT-US91-09160-20

  Query Match             4.2%;  Score 73.8;  DB 5;  Length 1378;
  Best Local Similarity   97.4%;  Pred. No. 1.1e-10;
  Matches   75;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     52 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 111

Qy     61 AATTCGGCACGAGCCGA 77
          |||||||||||||  ||
Db    112 AATTCGGCACGAGAAGA 128


RESULT 7
US-09-043-937A-3
; Sequence 3, Application US/09043937A
; Patent No. 6211432
;   GENERAL INFORMATION:
;        APPLICANT: BOUDET, ALAIN-MICHEL
;                   PICHON, MAGALIE
;                   GRIMA-PETTENATI, JACQUELINE
;                   BECKERT, MICHEL
;                   GAMAS, PASCAL
;                   BRIAT, JEAN-FRANCOIS
;        TITLE OF INVENTION: DNA SEQUENCES CODING FOR CINNAMOYL-CoA
;                            REDUCTASE, AND APPLICATIONS THEREOF IN THE CONTROL OF
;                            LIGNIN CONTENTS IN PLANTS
;        NUMBER OF SEQUENCES: 29
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: NIXON & VANDERHYE, P.C.
;             STREET: 110 NORTH GLEBE ROAD, 8TH FLOOR
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/043,937A
;             FILING DATE: 24-Jul-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: WO PCT/FR96/01544
;             FILING DATE: 03-OCT-1996
;             APPLICATION NUMBER: FR 95.11623
;             FILING DATE: 03-OCT-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: SADOFF, B.J.
;             REGISTRATION NUMBER: 36.663
;             REFERENCE/DOCKET NUMBER: 1487-20
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-816-4000
;             TELEFAX: 703-816-4100
;   INFORMATION FOR SEQ ID NO: 3:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1556 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA to mRNA
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  195..1310
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-043-937A-3

  Query Match             4.2%;  Score 73.8;  DB 4;  Length 1556;
  Best Local Similarity   97.4%;  Pred. No. 1.1e-10;
  Matches   75;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     52 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 111

Qy     61 AATTCGGCACGAGCCGA 77
          |||||||||||||  ||
Db    112 AATTCGGCACGAGAGGA 128


RESULT 8
US-09-043-937A-1
; Sequence 1, Application US/09043937A
; Patent No. 6211432
;   GENERAL INFORMATION:
;        APPLICANT: BOUDET, ALAIN-MICHEL
;                   PICHON, MAGALIE
;                   GRIMA-PETTENATI, JACQUELINE
;                   BECKERT, MICHEL
;                   GAMAS, PASCAL
;                   BRIAT, JEAN-FRANCOIS
;        TITLE OF INVENTION: DNA SEQUENCES CODING FOR CINNAMOYL-CoA
;                            REDUCTASE, AND APPLICATIONS THEREOF IN THE CONTROL OF
;                            LIGNIN CONTENTS IN PLANTS
;        NUMBER OF SEQUENCES: 29
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: NIXON & VANDERHYE, P.C.
;             STREET: 110 NORTH GLEBE ROAD, 8TH FLOOR
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
```

```
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/043,937A
;              FILING DATE: 24-Jul-1998
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: WO PCT/FR96/01544
;              FILING DATE: 03-OCT-1996
;              APPLICATION NUMBER: FR 95.11623
;              FILING DATE: 03-OCT-1995
;         ATTORNEY/AGENT INFORMATION:
;              NAME: SADOFF, B.J.
;              REGISTRATION NUMBER: 36.663
;              REFERENCE/DOCKET NUMBER: 1487-20
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 703-816-4000
;              TELEFAX: 703-816-4100
;    INFORMATION FOR SEQ ID NO: 1:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1568 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: double
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA to mRNA
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  278..1306
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-043-937A-1

  Query Match             4.2%;  Score 73.8;  DB 4;  Length 1568;
  Best Local Similarity   97.4%;  Pred. No. 1.1e-10;
  Matches   75;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 106

Qy      61 AATTCGGCACGAGCCGA 77
           |||||||||||||  ||
Db     107 AATTCGGCACGAGGGGA 123


RESULT 9
US-09-227-357-26
; Sequence 26, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P1
;  CURRENT APPLICATION NUMBER: US/09/227,357
;  CURRENT FILING DATE: 1999-01-08
;  EARLIER APPLICATION NUMBER: PCT/US98/13684
;  EARLIER FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: 60/051,926
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,793
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,925
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,929
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,803
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,732
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,931
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,932
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,916
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,930
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,918
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,920
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,733
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,795
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,919
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,928
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/055,722
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,723
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,948
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,949
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,953
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,950
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,947
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,964
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/056,360
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,684
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,984
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,954
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/058,785
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,664
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,660
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,661
;  EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 1543
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (69)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (717)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (899)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-26

  Query Match             4.1%;  Score 73.2;  DB 4;  Length 1543;
  Best Local Similarity   73.8%;  Pred. No. 1.7e-10;
  Matches   93;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;
```

```
Qy      11 CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCAC 70
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      99 CCACCGCGGTGGCGGCCGTTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCAC 158

Qy      71 GAGCCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCG 130
           ||||||| || | | ||      | | | |||| || ||  | |   |  | | |
Db     159 GAGCCGAACAGTAGGACATGTCATGGCATTTTTGCTCACCCTTGTTCCACTCCTCCCCAG 218

Qy     131 ACGTAG 136
           |||| |
Db     219 CCGTTG 224
```

RESULT 10
US-09-227-357-125
; Sequence 125, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: PZ010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (87)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-125

Query Match 4.1%; Score 73; DB 4; Length 545;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 176

Qy      61 AATTCGGCACGAG 73
           |||||||||||||
Db     177 AATTCGGCACGAG 189
```

RESULT 11
US-09-183-861-54
; Sequence 54, Application US/09183861
; Patent No. 6365165
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle

```
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/183,861
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  09/022,765
;       FILING DATE:  12-FEB-1998
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.420C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  54:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1585 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-183-861-54

  Query Match             4.1%;  Score 73;  DB 4;  Length 1585;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-10;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 62

Qy      61 AATTCGGCACGAG 73
           |||||||||||||
Db      63 AATTCGGCACGAG 75


RESULT 12
US-09-022-765-54
; Sequence 54, Application US/09022765
; Patent No. 6375955
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Campos-Neto, Antonio
;    APPLICANT:  Webb, John R.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Skeiky, Yasir A.W.
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,765
;      FILING DATE:  12-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1585 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-022-765-54

  Query Match             4.1%;  Score 73;  DB 4;  Length 1585;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-10;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 62

Qy      61 AATTCGGCACGAG 73
           |||||||||||||
Db      63 AATTCGGCACGAG 75


RESULT 13
US-08-730-771-1
; Sequence 1, Application US/08730771
; Patent No. 5780609
;  GENERAL INFORMATION:
;    APPLICANT:  Marshall, Lisa
;    APPLICANT:  Roshak, Amy
;    TITLE OF INVENTION:  A NOVEL HUMAN B-CELL SURFACE MOLECULE
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-2799
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/730,771
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  P50380
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2775 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
```

```
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-730-771-1

  Query Match             4.1%;  Score 73;  DB 1;  Length 2775;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-10;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 96

Qy      61 AATTCGGCACGAG 73
           |||||||||||||
Db      97 AATTCGGCACGAG 109


RESULT 14
US-09-060-208-1
; Sequence 1, Application US/09060208
; Patent No. 6221629
;  GENERAL INFORMATION:
;    APPLICANT:  Marshall, Lisa
;    APPLICANT:  Roshak, Amy
;    TITLE OF INVENTION:  THE DNA SEQUENCE OF HUMAN RP-105
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482-0980
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/060,208
;      FILING DATE:  April 15, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F.
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  P50380-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2775 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-09-060-208-1

  Query Match             4.1%;  Score 73;  DB 4;  Length 2775;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-10;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 96

Qy      61 AATTCGGCACGAG 73
           |||||||||||||
Db      97 AATTCGGCACGAG 109


RESULT 15
US-08-472-934-5
; Sequence 5, Application US/08472934
; Patent No. 5753446
;  GENERAL INFORMATION:
;    APPLICANT:  JOHNSON, GARY L.
;    TITLE OF INVENTION:  METHOD AND PRODUCT FOR REGULATING CELL
;    TITLE OF INVENTION:  RESPONSIVENESS TO EXTERNAL SIGNALS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive and Cockfield
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,934
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/440,421
;      FILING DATE:  15-May-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/354,516
;      FILING DATE:  21-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 5,405,941
;      FILING DATE:  15-Apr-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/323,460
;      FILING DATE:  14-Oct-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/11690
;      FILING DATE:  14-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04178
;      FILING DATE:  15-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Giulio A. DeConti, Jr. Esq.
;      REGISTRATION NUMBER:  31,503
;      REFERENCE/DOCKET NUMBER:  CPI-004DVCP2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3089 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  400..2280
US-08-472-934-5

  Query Match             4.1%;  Score 73;  DB 1;  Length 3089;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-10;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 60
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 AGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGG 70

Qy      61 AATTCGGCACGAG 73
           |||||||||||||
Db      71 AATTCGGCACGAG 83
```

Search completed: April 26, 2003, 02:28:24
Job time : 63 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 00:56:15 ; Search time 1751 Seconds
(without alignments)
16408.220 Million cell updates/sec

Title: US-09-854-300-7
Perfect score: 1774
Sequence: 1 agctggagctccaccgcggt.....ataggcaagtttcctctcag 1774

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1205.2	67.9	1921	11	AK018582	AK018582 Mus muscu
2	1202	67.8	2741	11	AK004847	AK004847 Mus muscu
3	836.8	47.2	1954	11	BC030951	BC030951 Homo sapi
4	680.4	38.4	1806	11	BC012931	BC012931 Homo sapi
5	673.8	38.0	811	13	BI600046	BI600046 603251431
6	654.2	36.9	1000	13	BI600419	BI600419 603246458
7	644.2	36.3	676	10	AV690200	AV690200 AV690200
8	641	36.1	773	13	BI601096	BI601096 603249588
9	623.4	35.1	840	11	AK008312	AK008312 Mus muscu
10	622.8	35.1	1150	13	BI488609	BI488609 603021137
11	622.4	35.1	906	12	BG165370	BG165370 602346005
12	622.2	35.1	683	10	AV686169	AV686169 AV686169
13	615.4	34.7	996	12	BG176206	BG176206 602335351
14	577	32.5	681	13	BI220592	BI220592 602936079
15	564.4	31.8	683	10	AV692456	AV692456 AV692456
16	542.8	30.6	557	13	BM507037	BM507037 ih24a05.y
17	537.2	30.3	904	12	BF664617	BF664617 602117901
18	531.2	29.9	596	12	BF231341	BF231341 253699 BA
19	524.2	29.5	862	12	BF668554	BF668554 602123601
20	515.6	29.1	902	12	BG165006	BG165006 602343868
21	513.8	29.0	841	12	BF696359	BF696359 602124911
22	509.8	28.7	812	13	BI110345	BI110345 602900777
23	502.2	28.3	872	12	BF699229	BF699229 602126878
24	499.4	28.2	789	12	BF669635	BF669635 602120364
25	496	28.0	879	12	BF666897	BF666897 602121530
26	492.4	27.8	759	12	BG400279	BG400279 602464672
27	482.4	27.2	774	12	BG432146	BG432146 602496225
28	481.6	27.1	841	12	BF666680	BF666680 602121368
29	481.6	27.1	847	12	BE958515	BE958515 601645138
30	475.6	26.8	804	13	BI653069	BI653069 603300779
31	471.8	26.6	856	12	BE958087	BE958087 601644929
32	469	26.4	680	10	BE308209	BE308209 601095109
33	468.6	26.4	884	12	BF542034	BF542034 602069189
34	466.4	26.3	835	12	BE958350	BE958350 601644838
35	461.4	26.0	622	12	BF137973	BF137973 601784651
36	459.2	25.9	725	13	BI557834	BI557834 603236825
37	458.6	25.9	466	12	BF382058	BF382058 601816390
38	456.4	25.7	732	12	BF246687	BF246687 601855477
39	443.6	25.0	813	13	BI648028	BI648028 603279210
c 40	442.6	24.9	619	13	BG991456	BG991456 MR1-HT106
41	441.8	24.9	889	12	BG172246	BG172246 602333390
42	440	24.8	862	12	BF700761	BF700761 602128561
43	435.4	24.5	854	12	BF670202	BF670202 602124352
44	431.4	24.3	759	12	BE958295	BE958295 601645073
45	427.2	24.1	942	12	BF967109	BF967109 602287635

ALIGNMENTS

RESULT 1
AK018582
LOCUS AK018582 1921 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130001F19:BRAIN CDNA, CLONE MNCB-3816, SIMILAR TO AF171875 G1-RELATED ZINC FINGER PROTEIN (MUS MUSCULUS), full insert sequence.
ACCESSION AK018582
VERSION AK018582.1 GI:12858361
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male cecum cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:9130001F19.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374

PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5 (bases 1 to 1921)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.2. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES Location/Qualifiers
source 1. .1921
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9130001F19"
/db_xref="MGD:MGI:1894834"
/db_xref="taxon:10090"
/clone="9130001F19"
/sex="male"
/tissue_type="cecum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 189. .1475
/note="BRAIN CDNA, CLONE MNCB-3816, SIMILAR TO AF171875 G1-RELATED ZINC FINGER PROTEIN (MUS MUSCULUS) data source:SPTR, source key:Q9JJF8, evidence:ISS putative"
/codon_start=1
/protein_id="BAB31291.1"
/db_xref="GI:12858362"
/db_xref="MGD:MGI:1914139"
/translation="MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWT AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNF TVPTVWGSTVQVSWLALIQRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSH PGAGDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFII TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDS CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG SVSLQVPVSNEASNTASPHEEDSRSETASSGYASVQGADEPPLEEHAQSANENLQLVN HEANSVAVDVVPHVDNPTFEEDETPDQEAAVREIKS"

BASE COUNT 466 a 466 c 503 g 486 t
ORIGIN

Query Match 67.9%; Score 1205.2; DB 11; Length 1921;
Best Local Similarity 89.6%; Pred. No. 1.1e-271;
Matches 1364; Conservative 0; Mismatches 149; Indels 10; Gaps 6;

```
Qy    252 GCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCT 311
          | | ||||||||||||||||||||| |||  |||||||| ||||||||| ||||||||
Db    178 GAGTGCGCGCCATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAG 237

Qy    312 TTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCC 371
          | || |  | ||||| |||||||||| ||||| |||||||||| || ||||||||||
Db    238 CTGCCCGGCTACTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCC 297

Qy    372 GGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGG 431
          |||| || ||||| ||||||||||||||||||||||||||||||||||||||||||||| |
Db    298 GGGGAGCCGAAGCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCG 357

Qy    432 GAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 491
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 GAGTGAACCGCACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 417

Qy    492 TGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTA 551
          ||||||| ||   | |||||||||||||||||||||||||||||||||||| ||||||||||
Db    418 TGGAGCCCGTCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTA 477

Qy    552 ACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGT 611
          |||||||||| |||||||||||||||||||||||||| ||||| |||||||| |||||||
Db    478 ACCCGCACACCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGT 537

Qy    612 TGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATG 671
          |||||||||||||||||| || |||||||||||||| ||||||||||||||||||||  |
Db    538 TGGCCCTCATCCAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAG 597

Qy    672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
          ||||| |||| |||||||| |||||||||||||||| ||||||||||||||||||||||
Db    598 AGAGAGGGGCTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCC 657
```

```
Qy    732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
          ||||||||||||||||||| | ||||||||||||||||||||| |||||||||||||| |
Db    658 CCATGTCTCACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAA 717

Qy    792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
          |||||||||||| |||||||||||||||||| ||||| |||||||||| ||||||||||
Db    718 CAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGA 777

Qy    852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTA 911
          |||||||||||||||||||||||| |||||||||| |||||||||||||||||||||| |
Db    778 AAAAACATGGCCCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCA 837

Qy    912 TTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATG 971
          | |||||||| ||||| |||||||||| |||||||||||||||||||| |||| ||||
Db    838 TAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATG 897

Qy    972 CAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAA 1031
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    898 CAAGAGCTCAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAA 957

Qy   1032 GGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTT 1091
          |||||| || |||||  |||||||||||||||||||||||||||||||||||||||||| |
Db    958 AGCTTCAGCTGCGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCT 1017

Qy   1092 GTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACC 1151
          ||||||||||||||| | ||||| |||||||||||||| |||||| |||| ||||| |
Db   1018 GTGCTGTGTGCATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATC 1077

Qy   1152 ATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGT 1211
          |||||||||||||||||||| ||||| |||||  || |||||| ||||||||||||||
Db   1078 ATATTTTCCATAAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGT 1137

Qy   1212 GCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1271
          |||| |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db   1138 GCAAGTGTGACATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1197

Qy   1272 CTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCCATGAAGAGGATA 1331
          |||||||||| ||||| || ||||||  |||||||| ||||||  |||||||||||||| |
Db   1198 CTTTACAAGTTCCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACA 1257

Qy   1332 ATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTC 1391
          |||||| ||||| |||||||||||||||||||||||||| ||| ||||||| || ||||
Db   1258 GTCGCAGTGAGACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTC 1317

Qy   1392 TGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATT 1451
          |||||||||| | ||||||| ||||||||| ||||||||||||||||||||||||||||||
Db   1318 TGGAGGAACATGCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATT 1377

Qy   1452 CTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTC 1511
          ||||||| ||||||||| |||| ||||| |||||||||||||||||||||| |||||||
Db   1378 CTGTGGCCGTGGATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTC 1437

Qy   1512 CTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTT-AAAATCTGTGTAAATAGAAAACT 1570
          || |||||||| | |||||||| || |||||||||.|||||||||| ||||||||||||
Db   1438 CTGATCAAGAGGCAGCTGTTCGGGAGATTAAATCTTAAAAATCTGTGTCAATAGAAAACT 1497

Qy   1571 TGAACCATTAG-TAATAACAGAACTGCCAATCAGGGCCTAGTTT-CTATTAATAAATTGG 1628
          |||||| |||| ||| ||||| |||||||||||||||||||| |||| ||| || |||
Db   1498 TGAACCGTTAGTTAACAACAGGACTGCCAATCAGGGCCTAGTTTACTATGAATGAACTGG 1557

Qy   1629 ATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATA 1688
          ||||     ||||| |||| |||||||||||||||| || | ||| ||||||| |||||
Db   1558 GTAAA---CGTAAAACAAGAATGATACTGAAAGTGCTGAGGTAACTTATATTATACTATA 1614

Qy   1689 GTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATA 1748
          ||||||   | | || ||||||| ||   || | |||| |  |||||||  ||||
Db   1615 GTTAAA--TGGCTTAACATATTTACCCCAGTACCGTTTT--CCACAAACTCACCATAACG 1670

Qy   1749 TTTTTCATAGGCAAGTTTCCTCT 1771
          |||||||||||||||||||||||
Db   1671 TTTTTCATAGGCAAGTTTCCTCT 1693
```

RESULT 2
AK004847

LOCUS	AK004847 2741 bp mRNA linear HTC 19-JAN-2002
DEFINITION	Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300002C13:BRAIN CDNA, CLONE MNCB-3816, SIMILAR TO AF171875 G1-RELATED ZINC FINGER PROTEIN (MUS MUSCULUS), full insert sequence.
ACCESSION	AK004847
VERSION	AK004847.1 GI:12836339
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1300002C13.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5 (bases 1 to 2741)
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,

Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES Location/Qualifiers

source 1. .2741
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1300002C13"
/db_xref="MGD:MGI:1895035"
/db_xref="taxon:10090"
/clone="1300002C13"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

CDS 188. .1474
/note="BRAIN CDNA, CLONE MNCB-3816, SIMILAR TO AF171875 G1-RELATED ZINC FINGER PROTEIN (MUS MUSCULUS)
data source:SPTR, source key:Q9JJF8, evidence:ISS putative"
/codon_start=1
/protein_id="BAB23613.1"
/db_xref="GI:12836340"
/db_xref="MGD:MGI:1914139"
/translation="MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWT AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNF TVPTVWGSTVQVSWLALIQRGGSCTFADNIHLASERGASGAVIFNFPGTRNEVIPMSH PGAGDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFII TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDS CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDG SVSLQVPVSNEASNTASPHEEDSRSETASSGYASVQGADEPPLEEHAQSANENLQLVN HEANSVAVDVVPHVDNPTFEEDETPDQEAAVREIKS"

polyA_signal 2725. .2730
/note="putative"

polyA_site 2741
/note="putative"

BASE COUNT 764 a 578 c 643 g 756 t

ORIGIN

Query Match 67.8%; Score 1202; DB 11; Length 2741;
Best Local Similarity 89.4%; Pred. No. 6.2e-271;
Matches 1362; Conservative 0; Mismatches 151; Indels 10; Gaps 6;

```
Qy    252 GCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCT 311
Db    177 GAGTGCGCGCCATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAG 236
Qy    312 TTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCC 371
Db    237 CTGCCCGGCTACTGGCTTGGTGCTTCCTTCTGGCTCTGAGTCCGCACGCGCCCGGTTCCC 296
Qy    372 GGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGG 431
Db    297 GGGGAGCCGAAGCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCG 356
Qy    432 GAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 491
Db    357 GAGTGAACCGCACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 416
Qy    492 TGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTA 551
Db    417 TGGAGCCCGTCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTA 476
Qy    552 ACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGT 611
Db    477 ACCCGCACACCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGT 536
Qy    612 TGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATG 671
Db    537 TGGCCCTCATCCAACGCGGTGGAAGCTGCACCTTCGCGGACAATATCCATCTGGCTTCAG 596
Qy    672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
Db    597 AGAGAGGGGCTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCC 656
Qy    732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
Db    657 CCATGTCTCACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAA 716
Qy    792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
Db    717 CAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGA 776
Qy    852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTA 911
Db    777 AAAAACATGGCCCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCA 836
Qy    912 TTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATG 971
Db    837 TAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATG 896
Qy    972 CAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAA 1031
Db    897 CAAGAGCTCAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAA 956
Qy   1032 GGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTT 1091
Db    957 AGCTTCAGCTGCGCACCTTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGCT 1016
Qy   1092 GTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCTTAACGTGCAACC 1151
Db   1017 GTGCTGTGTGCATTGAGCTCTATAAGCCAAATGATTTGGTGCGCATCCTAACCTGCAATC 1076
Qy   1152 ATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGT 1211
Db   1077 ATATTTTCCATAAGACATGTGTGGACCCGTGGCTTTTAGAACACAGGACTTGCCCCATGT 1136
Qy   1212 GCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1271
Db   1137 GCAAGTGTGACATTCTCAAAGCTCTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGT 1196
Qy   1272 CTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATA 1331
Db   1197 CTTTACAAGTTCCTGTTTCTAATGAAGCATCTAATACTGCCTCTCCCCATGAAGAGGACA 1256
Qy   1332 ATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTC 1391
Db   1257 GTCGCAGTGAGACTGCATCATCTGGATATGCTTCAGTACAAGGAGCAGATGAGCCACCTC 1316
```

```
Qy   1392 TGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATT 1451
          |||||||||| | ||||||| ||||||||| ||||||||||||||||||||||||||||||
Db   1317 TGGAGGAACATGCGCAGTCAGCAAATGAAAATCTACAGCTGGTAAACCATGAAGCAAATT 1376

Qy   1452 CTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTC 1511
          ||||||| ||||||||| |||| ||||| |||||||||||||||||||||| |||||||
Db   1377 CTGTGGCCGTGGATGTTGTTCCCCATGTTGACAACCCAACCTTTGAAGAAGATGAAACTC 1436

Qy   1512 CTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTT-AAAATCTGTGTAAATAGAAAACT 1570
          || |||||||| | |||||||| || |||||||||| |||||||||| ||||||||||
Db   1437 CTGATCAAGAGGCAGCTGTTCGGGAGATTAAATCTTAAAAATCTGTGTCAATAGAAAACT 1496

Qy   1571 TGAACCATTAG-TAATAACAGAACTGCCAATCAGGGCCTAGTTT-CTATTAATAAATTGG 1628
          |||||| |||| ||| ||||| ||||||||||||||||||||| |||| ||| || |||
Db   1497 TGAACCGTTAGTTAACAACAGGACTGCCAATCAGGGCCTAGTTTACTATGAATGAACTGG 1556

Qy   1629 ATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATATTATGCTATA 1688
            ||||    ||||| |||| |||||||||||||| || | ||| ||||||| |||||
Db   1557 GTAAA---CGTAAAACAAGAATGATACTGAAAGTGCTGAGGTAACTTATATTATACTATA 1613

Qy   1689 GTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTCATTATAATA 1748
          ||||||   | | || |||||||| ||   || | |||| | |||||||| ||||
Db   1614 GTTAAA--TGGCTTAACATATTTACCCCAGTACCGTTTT--CCACAAACTCACCATAACG 1669

Qy   1749 TTTTTCATAGGCAAGTTTCCTCT 1771
          |||||||||||||||||||||||
Db   1670 TTTTTCATAGGCAAGTTTCCTCT 1692
```

RESULT 3
BC030951
LOCUS BC030951 1954 bp mRNA linear HTC 13-JUN-2002
DEFINITION Homo sapiens, hypothetical protein FLJ23516, clone IMAGE:4280352, mRNA.
ACCESSION BC030951
VERSION BC030951.1 GI:21410659
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1954)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 41 Row: c Column: 10
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13375696
This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers
source 1. .1954
/organism="Homo sapiens"
/db_xref="LocusID:79589"
/db_xref="taxon:9606"
/clone="IMAGE:4280352"
/tissue_type="Brain, primitive neuroectodermal"
/clone_lib="NIH_MGC_56"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

BASE COUNT 676 a 294 c 346 g 638 t
ORIGIN

Query Match 47.2%; Score 836.8; DB 11; Length 1954;
Best Local Similarity 98.7%; Pred. No. 2.3e-185;
Matches 864; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

```
Qy    900 TGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAA 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAA 60

Qy    960 GGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAA 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAA 120

Qy   1020 AAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTG 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTG 180

Qy   1080 ATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATCT 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    181 ATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTACGCATTT 240

Qy   1140 TAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAACACAAGA 1199
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| ||
Db    241 TAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAGAACACAGGA 300

Qy   1200 CTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAG 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGATGTTGAAG 360

Qy   1260 ATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCC 1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCCTCCTCCC 420

Qy   1320 ATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAG 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAGGGAACAG 480

Qy   1380 ATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACC 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTGGTAAACC 540

Qy   1440 ATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGACAACCCAACCTTTGAAG 1499
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    541 ATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTTGACAACCCAACCTTTGAAG 600

Qy   1500 AAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTA 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAAATCTGTGTA 660

Qy   1560 AATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTA 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AATAGAAAACTTGAACCATTAGTAATAACAGAACTGCCAATCAGGGCCTAGTTTCTATTA 720

Qy   1620 ATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATAT 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATAAATTGGATAAATTTAATAAAATAAGAGTGATACTGAAAGTGCTCAGATGACTAATAT 780

Qy   1680 TATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTGTTAACTTTTTTCCACCAAACTC 1739
          |||||||||||||||   | | ||||||||||||||||||||||||||||| |||||||
Db    781 TATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTGTTAACTTTTTTCCA-CAAACTC 837

Qy   1740 ATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          |||||||||||||||||||||||||||||||||
Db    838 ATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 872
```

```
RESULT 4
BC012931
LOCUS       BC012931                1806 bp    mRNA    linear   HTC 22-AUG-2001
DEFINITION  Homo sapiens, Similar to hypothetical protein FLJ23516, clone
            IMAGE:4456033, mRNA.
ACCESSION   BC012931
VERSION     BC012931.1  GI:15277870
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1806)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 28 Row: b Column: 9
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 10440232
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1. .1806
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4456033"
                     /tissue_type="Kidney, hypernephroma"
                     /clone_lib="NIH_MGC_89"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      616 a    266 c    317 g    607 t
ORIGIN

  Query Match             38.4%;  Score 680.4;  DB 11;  Length 1806;
  Best Local Similarity   98.6%;  Pred. No. 1e-148;
  Matches  707;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    2;

Qy   1058 GGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAA 1117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAA 60

Qy   1118 CCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGAC 1177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGAC 120

Qy   1178 CCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTG 1237
          |||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||
Db    121 CCATGGCTGTTAGAACACAGGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTG 180

Qy   1238 GGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAA 1297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAA 240

Qy   1298 ATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGA 1357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGA 300

Qy   1358 TATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAAT 1417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAAT 360

Qy   1418 GAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCAT 1477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCAT 420

Qy   1478 GTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAA 1537
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTTGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAA 480

Qy   1538 ATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCC 1597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCC 540

Qy   1598 AATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTG 1657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTG 600

Qy   1658 AAAGTGCTCAGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTTAACCTG 1717
          ||||||||||||||||||||||||||||||||||||    |  | ||||||||||||||||
Db    601 AAAGTGCTCAGATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTTAACCTG 658

Qy   1718 TTAACTTTTTTCCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 1774
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    659 TTAACTTTTTTCCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCTCAG 714


RESULT 5
BI600046
LOCUS       BI600046                811 bp    mRNA    linear   EST 07-SEP-2001
DEFINITION  603251431F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5303239 5',
            mRNA sequence.
ACCESSION   BI600046
VERSION     BI600046.1  GI:15492985
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11767  row: n  column: 08
            High quality sequence stop: 776.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5303239"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
```

for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 153 a 244 c 264 g 150 t
ORIGIN

Query Match 38.0%; Score 673.8; DB 13; Length 811;
Best Local Similarity 94.2%; Pred. No. 3.3e-147;
Matches 745; Conservative 0; Mismatches 37; Indels 9; Gaps 4;

```
Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      22 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTTCCTCCTGGCTCCGACG 81

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 141

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           |||| |||||||||||||||||||||||||H||||||||||||||||||||||||||||||
Db     142 ACCA-GGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 200

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 260

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     261 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTACCGG 320

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 380

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 440

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 500

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     501 CCGCACACGAATTTCACGGTGCCCACGATTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 560

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 620

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAA--CTTCCCCGGGACCCGCAATGAG----GTC 727
           ||| |||||||||||||||||||||||||    |||||||| ||||||||||    |||
Db     621 AGAGGGGCGTCTGGAGCCGTCATCTTTAAACTGTCCCCGGGAGCCGCAATGAGAGTCAAT 680

Qy     728 ATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATC--ATGATCGGCAATCTGA 785
             |||  |||||||||||||||||||||    ||||||||||    |||||||||
Db     681 CCCCCATTGTCTCACCCGGGTGCAGTAGAACATGTTGCAATCCATGAATCGGCAATTCTG 740

Qy     786 AAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAG 845
           || |  ||||||||||||||||||| |||||||||||||||||||| ||||||| | ||
Db     741 AAAGGCCAAAAATTCTGCAATCTATGCAAAGAGGCATACAAGTGACCATGGTCAAACAAA 800

Qy     846 TAGGGAAAAAA 856
             ||  ||||
Db     801 GTAGGGGAAAA 811
```

RESULT 6
BI600419
LOCUS BI600419 1000 bp mRNA linear EST 07-SEP-2001
DEFINITION 603246458F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5288774 5', mRNA sequence.
ACCESSION BI600419
VERSION BI600419.1 GI:15493358
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1000)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11730 row: c column: 15
High quality sequence stop: 701.
FEATURES Location/Qualifiers
source 1. .1000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5288774"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 206 a 299 c 321 g 174 t
ORIGIN

Query Match 36.9%; Score 654.2; DB 13; Length 1000;
Best Local Similarity 93.7%; Pred. No. 1.3e-142;
Matches 747; Conservative 0; Mismatches 43; Indels 7; Gaps 6;

```
Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      22 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTTCCTCCTGGCTCCGACG 81

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 141

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 ACCA-GGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCA-GGGCGCTAGGGAACTGCGGAGC 199

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 259

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     260 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 319

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 379

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    380 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 439

Qy    494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 499

Qy    554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGG-TT 612
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||
Db    500 CCGCACACGAATTTCACGGTGCCCACGATTTGGGGAAGCACCGTGCAAGTCTCTTGGTTT 559

Qy    613 GGCCCTCATCCAACGCGGCGGGGGCTGC-ACCTTCGCAGACAAGATCCATCTGGCTTATG 671
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    560 GGCCCTCATCCAACGCGGCGGGGGCTGCAACCTTCGCAGACAAGATCCATCTGGCTTATG 619

Qy    672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    620 AGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCAT-C 678

Qy    732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    679 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCCATGATCCGGCCATCTGAAGGG 738

Qy    792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
          | |||| |||| |||||||||| ||       |||||| |  |   |||||| ||||
Db    739 CCAAAAATCTGGAATCTATTCAAGAAG--CATACAGTGACCACTGGTCTAGAAGTCGGGA 796

Qy    852 AAAAACATGGCCCTTGG 868
          | ||||||||||   ||
Db    797 ACAAACATGGCCTGGGG 813
```

```
RESULT 7
AV690200
LOCUS       AV690200                676 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV690200 GKC Homo sapiens cDNA clone GKCAKA02 5', mRNA sequence.
ACCESSION   AV690200
VERSION     AV690200.1  GI:10292063
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GKCAKA02"
                     /clone_lib="GKC"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       94 a    223 c    231 g    125 t      3 others
ORIGIN
```

```
  Query Match             36.3%;  Score 644.2;  DB 10;  Length 676;
  Best Local Similarity   98.1%;  Pred. No. 2.8e-140;
  Matches  661;  Conservative    0;  Mismatches   11;  Indels    2;  Gaps    1;

Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 63

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 123

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db.    124 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 183

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 243

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 303

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 363

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 423

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||  |||||||||||| ||||||||||||||||||||
Db     424 GAGCCTGTGGCTGGGGTCCTGGTACCCGCCGACGGGCCCGGNGCGCTTAACGCCTGTAAC 483

Qy     554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 543

Qy     614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     544 GCCCTCATCCAACGCGGCGGNGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 603

Qy     674 AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCC 733
           ||| ||||||||||||||||||||  ||||||||||||    |||||||||| |||
Db     604 AGAGGGGCGTCTGGAGCCGTCATCTTT--CTTCCCCGGGACCCCGCATGAGGTCATNCCC 661

Qy     734 ATGTCTCACCCGGG 747
           ||||||| || |||
Db     662 ATGTCTCCCCGGGG 675
```

```
RESULT 8
BI601096
LOCUS       BI601096                773 bp    mRNA    linear   EST 07-SEP-2001
DEFINITION  603249588F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5301380 5',
            mRNA sequence.
ACCESSION   BI601096
VERSION     BI601096.1  GI:15494035
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
```

```
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
           Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11762  row: p  column: 21
           High quality sequence stop: 759.
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5301380"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length  clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      130 a    242 c    254 g    146 t      1 others
ORIGIN

  Query Match             36.1%;  Score 641;  DB 13;  Length 773;
  Best Local Similarity   96.3%;  Pred. No. 1.6e-139;
  Matches  720;  Conservative    0;  Mismatches   21;  Indels    7;  Gaps    6;

Qy      74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      22 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTTCCTCCTGGCTCCGACG 81

Qy     134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 141

Qy     194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
           |||||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db     142 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCA-GGGCGCTAGGGGACTGCGGAGC 200

Qy     254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 260

Qy     314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 320

Qy     374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 380

Qy     434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 440

Qy     494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 500

Qy     554 CCGCACACGAATTTCAC-GGTGCCCACGGTTTGGGGAAGC-ACCGTGCAAGTCTCTTGG- 610
           ||||||||||||||||| |||||||||| |||||||||| |||||||||||||||||||
Db     501 CCGCACACGAATTTCACGGGTGCCCACGANTTGGGGAAGCAACCGTGCAAGTCTCTTGGT 560

Qy     611 TTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTAT 670
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 TTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTAT 620

Qy     671 GAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCC--GCAATGAGGTCA 728
           |||||| |||||||||||||||||||||||||||||||||||||||    |||||||||
Db     621 GAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCCGCAAATGAGGTCA 680

Qy     729 TCCCCATGTCTCACCCGGGTGCAGTAGACATT-GTTGCAATCATGATCGGCAATCTGAAA 787
           |||||||||||||||||||||||||||||||| |||||||||||||  |  |     |
Db     681 TCCCCATGTCTCACCCGGGTGCAGTAGACATTGGTTGCAATCATGATTCGGCAATCTGCA 740

Qy     788 GGCACAAAAATTCTGCAATCTATTCAAA 815
           |  |||||||||||||||||||| ||
Db     741 AGGCCAAAAATTCTGCAATCTATTCCAA 768


RESULT 9
AK008312
LOCUS       AK008312                 840 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010104G21:BRAIN CDNA, CLONE MNCB-3816,
            SIMILAR TO AF171875 G1-RELATED ZINC FINGER PROTEIN (MUS MUSCULUS),
            full insert sequence.
ACCESSION   AK008312
VERSION     AK008312.1  GI:12842420
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
            mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2010104G21.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
```

```
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5  (bases 1 to 840)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2010104G21"
                     /db_xref="MGD:MGI:1895748"
                     /db_xref="taxon:10090"
                     /clone="2010104G21"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             131. .>840
                     /note="BRAIN CDNA, CLONE MNCB-3816, SIMILAR TO AF171875
                     G1-RELATED ZINC FINGER PROTEIN (MUS MUSCULUS)
                     data source:SPTR, source key:Q9JJF8, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB25595.1"
                     /db_xref="GI:12842421"
                     /db_xref="MGD:MGI:1914139"
                     /translation="MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWT
                     AYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNF
                     TVPTVWGSTVQVSWLALIQRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSH
                     PGAGDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSPFII
                     TAATVGYFIFYSARRLRNA"
BASE COUNT      145 a    259 c    259 g    177 t
ORIGIN

  Query Match             35.1%;  Score 623.4;  DB 11;  Length 840;
  Best Local Similarity   91.5%;  Pred. No. 2.2e-135;
  Matches  660;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     252 GCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCT 311
           | | ||||||||||||||||||||| ||| |||||||| ||||||||| ||||||||
Db     120 GAGTGCGCGCCATGGGGCCGCCGCCCGGGATCGGGGTCTACTGCCGCGGCGGCTGCGGAG 179

Qy     312 TTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCC 371
           | || |  | ||||| |||||||||| |||||||||||||||| || ||||||||||
Db     180 CTGCCCGGCTACTGGCTTGGTGCTTCCTTCTGGCCCTGAGTCCGCACGCGCCCGGTTCCC 239

Qy     372 GGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGG 431
           |||| || ||||| ||||||||||||||||||||||||||||||||||||||||||||| |
Db     240 GGGGAGCCGAAGCCGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACCG 299

Qy     432 GAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 491
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GAGTGAACCGCACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGC 359

Qy     492 TGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTA 551
           ||||||| ||  | |||||||||||||||||||||||||||||||||||| ||||||||||
Db     360 TGGAGCCCGTCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTA 419

Qy     552 ACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGT 611
           |||||||||| |||||||||||||||||||||||||||||||||| |||| |||||||| |||||||
Db     420 ACCCGCACACCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGT 479

Qy     612 TGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATG 671
           |||||||||||||||||| || |||||||||||||| |||||||||||||||||||  |
Db     480 TGGCCCTCATCCAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAG 539

Qy     672 AGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCC 731
           ||||| |||| |||||||| |||||||||||||||| ||||||||||||||||||||||
Db     540 AGAGAGGGGCTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCC 599

Qy     732 CCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCA 791
           |||||||||||||||||| | ||||||||||||||||||| |||||||||||||| |
Db-    600 CCATGTCTCACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAA 659

Qy     792 CAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGA 851
           |||||||||||| |||||||||||| ||||| |||||||||| |||||||||||||||||
Db     660 CAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGGA 719

Qy     852 AAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTCGTTTCTGTGTCCTTTTTTA 911
           |||||||||||||||||||||||||| |||||||||| |||||||||||||||||||| |
Db     720 AAAAACATGGCCCTTGGGTGAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCA 779

Qy     912 TTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATG 971
           | |||||||| |||| |||||||||| |||||||||||||||||||  |||| ||||
Db     780 TAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATG 839

Qy     972 C 972
           |
Db     840 C 840


RESULT 10
BI488609
LOCUS       BI488609                1150 bp    mRNA    linear   EST 28-AUG-2001
DEFINITION  603021137F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191785 5',
            mRNA sequence.
ACCESSION   BI488609
VERSION     BI488609.1  GI:15327837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11479  row: j  column: 10
            High quality sequence stop: 740.
FEATURES             Location/Qualifiers
```

```
     source          1. .1150
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5191785"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning). Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      264 a    315 c    336 g    235 t
ORIGIN

  Query Match             35.1%;  Score 622.8;  DB 13;  Length 1150;
  Best Local Similarity   95.8%;  Pred. No. 3.1e-135;
  Matches  747;  Conservative    0;  Mismatches   22;  Indels   11;  Gaps   10;

Qy    174 CTGGCGCGCACCTGCTCAAGACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGG 233
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||| ||
Db      1 CTGGCGCGCACCTGCTCAAGACCA-GGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCA-GG 58

Qy    234 GCGCTAGGGAACTGCGGAGCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCT 293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59 GCGCTAGGGAACTGCGGAGCGCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCT 118

Qy    294 GCCGCGGTGGCTGCGGCTTTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTC 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 GCCGCGGTGGCTGCGGCTTTTCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTC 178

Qy    354 CGCAGGCACCCGGTTCCCG-GGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCC 412
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    179 CGCAGGCACCCGGTTCCCGTGGGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCC 238

Qy    413 TGGCGGGTTCCGCACACGGGAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTG 472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 TGGCGGGTTCCGCACACGGGAGTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTG 298

Qy    473 TACGGCCAGGACTCGCCGCTGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCC 532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 TACGGCCAGGACTCGCCGCTGGAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCC 358

Qy    533 GGGGCGCTTAAC-GCCTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAG 591
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    359 GGGGCGCTTAACTGCCTGTAACCCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAG 418

Qy    592 CACCGTGCAAGTCTCTTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGA 651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CACCGTGCAAGTCTCTTGGTTGGCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGA 478

Qy    652 C-AAGATCCATCTGGCTTATGAGAGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCG 710
          | ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    479 CGAAGATCCATCTGGCTTATGAGAGAGGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCG 538

Qy    711 GGACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCA 770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 GGACCCGCAATGAGGTCATCCCCATGTCTCACCCGGGTGCAGTAGACATTGTTGCAATCA 598

Qy    771 TGATCGGCAATCTGAAAGGCAC-AAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTG 829
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    599 TGATCGGCAATCTGAAAGGCACAAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTG 658

Qy    830 ACAATGGTCATAGAAGTA-GGGAAAAAACATGGCCCTTGGGTGAATC-ACTATTCAATTT 887
          |||||||||||||||||| |||||||||||||||| |  |||||||| ||||||||| |
Db    659 ACAATGGTCATAGAAGTACGGGAAAAAACATGGCCTTGCGGTGAATCAACTATTCAAATA 718

Qy    888 TTTTCGTTTC--TGTGTCCTTTTTTATTA-TTACGGCGGCAACTGTGGGCTATTTTATCT 944
          |||||||  |   |||||||||    |||| ||||||||||||||    |      ||| | |
Db    719 TTTTCGTAACCTGGTGTCCTTTCCCATTACTTACGGCGGCAACTTGTGCGCTATTTGTAT 778
```

```
RESULT 11
BG165370
LOCUS       BG165370                906 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602346005F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4456033 5',
            mRNA sequence.
ACCESSION   BG165370
VERSION     BG165370.1  GI:12672073
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov).
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10250  row: j  column: 02
            High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4456033"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      275 a    204 c    183 g    244 t
ORIGIN

  Query Match             35.1%;  Score 622.4;  DB 12;  Length 906;
  Best Local Similarity   96.0%;  Pred. No. 3.7e-135;
  Matches  692;  Conservative    0;  Mismatches   22;  Indels    7;  Gaps    5;

Qy   1058 GGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAA 1117
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      1 GGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAA-TGTATAAA 59

Qy   1118 CCAAATGATTTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGTGAC 1177
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 CCAAATGA-TTGGTACGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGTGAC 118

Qy   1178 CCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTG 1237
          |||||||||||| |||||| ||||||||||||||||||||||||||||||||||| |||
Db    119 CCATGGCTGTTAGAACACAGGACTTGCCCCATGTGCAAATGTGACATACTCAAAGC-TTG 177

Qy   1238 GGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAA 1297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAA 237

Qy   1298 ATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGA 1357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 ATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGA 297

Qy   1358 TATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAAT 1417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 TATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAAT 357
```

```
Qy   1418 GAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCAT 1477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 GAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCAT 417

Qy   1478 GTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAA 1537
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 GTTGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAA 477

Qy   1538 ATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCC 1597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 ATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGAACTGCC 537

Qy   1598 AATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTGATACTG 1657
          |||||||||||||||||||||| |||||  |   || ||||||||||||||||||||||||
Db    538 AATCAGGGCCTAGTTTCTATTAATAAATGGATAAATTTTAATAAAATAAGAGTGATACTG 597

Qy   1658 AAAGTGCTCAGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTT--AACC 1715
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||  | |
Db    598 AAAGTGCTCAGATGACTAATATTATGCTATAGTTAAACTGGCTTAAAAATATTTTAACGC 657

Qy   1716 TGTTAACTTTTTTCCACCAAACTCATTATAATATTTTTCA--TAGGCAAGTTTCCTCTCA 1773
          ||| ||  ||||| |  |||||||||||| |||||| ||    ||| ||||||||||||
Db    658 TGTCAAACTTTTTTCCACAAACTCATTATCATATTTTGCACCAGGGCCAGTTTCCTCTCA 717

Qy   1774 G 1774
          |
Db    718 G 718
```

RESULT 12
AV686169
LOCUS AV686169 683 bp mRNA linear EST 16-JAN-2002
DEFINITION AV686169 GKC Homo sapiens cDNA clone GKCAIE10 5', mRNA sequence.
ACCESSION AV686169
VERSION AV686169.1 GI:10288032
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 683)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCAIE10"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 98 a 222 c 228 g 135 t
ORIGIN

Query Match 35.1%; Score 622.2; DB 10; Length 683;
Best Local Similarity 97.0%; Pred. No. 4e-135;
Matches 655; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

```
Qy     74 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4 CCGAGGAGCTGCATCTGCGGCAACCTGTGTGCTGACGCTACGTGCCTCCTGGCTCCGACG 63

Qy    134 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 TAGCTCGCAGCTCCCCAGTCTCACTCCATTCCTTCCCCACCTGGCGCGCACCTGCTCAAG 123

Qy    194 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 ACCAGGGTCCTGCCAAGCGCTAGGAGGGCGCGTGCCAGGGGCGCTAGGGAACTGCGGAGC 183

Qy    254 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 GCGCGCGCCATGGGGCCGCCGCCTGGGGCCGGGGTCTCCTGCCGCGGTGGCTGCGGCTTT 243

Qy    314 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 TCCAGATTGCTGGCATGGTGCTTCCTGCTGGCCCTGAGTCCGCAGGCACCCGGTTCCCGG 303

Qy    374 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GGGGCTGAAGCAGTGTGGACCGCGTACCTCAACGTGTCCTGGCGGGTTCCGCACACGGGA 363

Qy    434 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 GTGAACCGTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTG 423

Qy    494 GAGCCTGTGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAAC 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 GAGCCTGTGGCTGGGGTCCTGGTACCCGCCGACGGGCCCGGTGCGCTTAACGCCTGTAAC 483

Qy    554 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTG 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 CCGCACACGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTTGTTG 543

Qy    614 GCCCTCATCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAG 673
          |||||||| |||||| |||||||||||||||||||||||| |||| |||| |||||||||
Db    544 GCCCTCATTCAACGCCGCGGGGGCTGCACCTTCGCAGACCAGATACATCTTGCTTATGAG 603

Qy    674 -AGATGGGCGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCC 732
           ||| |||||||||||||| ||||||| |||| |||||| ||| |||||||||||| ||
Db    604 AAGAAGGGCGTCTGGAGCCCGCATCTTTTACTTTCCCGGG-CCCCCAATGAGGTCATTCC 662

Qy    733 CATGTCTCACCCGGG 747
          ||||| | |||||||
Db    663 CATGTTTTACCCGGG 677
```

RESULT 13
BG176206
LOCUS BG176206 996 bp mRNA linear EST 06-FEB-2001
DEFINITION 6023353S1F2 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4458252 5', mRNA sequence.
ACCESSION BG176206
VERSION BG176206.1 GI:12682909
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10256 row: f column: 13
High quality sequence stop: 668.

FEATURES Location/Qualifiers
source 1. .996
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4458252"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 294 a 214 c 243 g 244 t 1 others
ORIGIN

Query Match 34.7%; Score 615.4; DB 12; Length 996;
Best Local Similarity 87.7%; Pred. No. 1.7e-133;
Matches 717; Conservative 0; Mismatches 97; Indels 4; Gaps 4;

```
Qy    773 ATCGGCAATCTGAAAGGCACAAAAATTCTGCAATCTATTCAAAGAGGCATACAAGTGACA 832
          || |||||||||||||| |||||||||||||| |||||||||||||||| ||||| |||
Db      1 ATTGGCAATCTGAAAGGAACAAAAATTCTGCAGTCTATTCAAAGAGGCATCCAAGTCACA 60

Qy    833 ATGGTCATAGAAGTAGGGAAAAAACATGGCCCTTGGGTGAATCACTATTCAATTTTTTTC 892
          |||||||| |||||||||||||||||||||| |||||||||| |||||||||| |||
Db     61 ATGGTCATCGAAGTAGGGAAAAAACATGGCCC-TGGGTGAATCATTATTCAATTTTCTTC 119

Qy    893 GTTTCTGTGTCCTTTTTTATTATTACGGCGGCAACTGTGGGCTATTTTATCTTTTATTCT 952
          |||||||||||||||| || |||||||| ||||| |||||||||| ||||||||||||
Db    120 GTTTCTGTGTCCTTTTTCATAATTACGGCAGCAACCGTGGGCTATTTCATCTTTTATTCT 179

Qy    953 GCTCGAAGGCTACGGAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAATTAAAGGCAGAT 1012
          ||||||||  |||| ||||||||||||||||||||||||||||||||| ||||||||||||
Db    180 GCTCGAAGATTACGAAATGCAAGAGCTCAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGAT 239

Qy   1013 GCTAAAAAAAGCTATTGGAAGGCTTCAACTACGCACACTGAAACAAGGAGACAAGGAAATT 1072
          |||||||||||||||||| |||||| || ||||| ||||||||||||||||||||||||||
Db    240 GCTAAAAAAGCTATTGGAAAGCTTCAGCTGCGCACCTTGAAACAAGGAGACAAGGAAATT 299

Qy   1073 GGCCCTGATGGAGATAGTTGTGCTGTGTGCATTGAATTGTATAAACCAAATGATTTGGTA 1132
          |||||||||||||||| |||||||||||||||  | ||||| ||||||||||||||
Db    300 GGCCCTGATGGAGATAGCTGTGCTGTGTGCATTGAGCTCTATAAGCCAAATGATTTGGTG 359

Qy   1133 CGCATCTTAACGTGCAACCATATTTTCCATAAGACATGTGTTGACCCATGGCTGTTAAAA 1192
          ||||||  |||| ||||| |||||||||||||||||||||||| ||||| ||||| ||| ||
Db    360 CGCATCCTAACCTGCAATCATATTTTCCATAAGACATGTGTGGACCCGTGGCTTTTAGAA 419

Qy   1193 CACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAAGCTTTGGGAATTGAGGTGGAT 1252
          |||| |||||||||||||||| |||||||| |||||||| ||||||||||||||||||||
Db    420 CACAGGACTTGCCCCATGTGCAAGTGTGACATTCTCAAAGCTCTGGGAATTGAGGTGGAT 479

Qy   1253 GTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCCAATGAAATATCTAATAGTGCC 1312
          |||||||||||||||||||||||||||||| ||||| || ||||||  |||||||| ||||
Db    480 GTTGAAGATGGATCAGTGTCTTTACAAGTTCCTGTTTCTAATGAAGCATCTAATACTGCC 539

Qy   1313 TCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCATCTGGATATGCTTCAGTACAG 1372
          ||  |||||||||||| | |||||| ||||| ||||||||||||||||||||||||||||
Db    540 TCTCCCCATGAAGAGGACAGTCGCAGTGAGACTGCATCATCTGGATATGCTTCAGTACAA 599

Qy   1373 GGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCAACAAATGAAAGTCTACAGCTG 1432
          ||| ||||||| || |||||||||||||| | ||||||| |||||||||| ||||||||||
Db    600 GGAGCAGATGAGCCACCTCTGGAGGAACATGCGCAGTCAGCAAATGAAAATCTACAGCTG 659

Qy   1433 GTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATTCCTCATGTNGAC-AACCCAAC 1491
          || |||||||||||||||||||||     |-  | | || ||||| ||| ||||||||
```

```
Db    660 GT-AACCATGAAGCAAATTCTGTGGGCCGTGGATGTGTCCCCCATGTTGACAAACCCAAC 718

Qy   1492 CTTTGAAG-AAGACGAAACTCCTAATCAAGAGACTGCTGTTCGAGAAATTAAATCTTAAA 1550
          |||||||| |||| ||||  ||| |||||||| | | ||| || || ||| | ||||| |
Db    719 CTTTGAAGAAAGATGAAATCCCTGATCAAGAGGCAGTTGTCCGGGAGATTCAGTCTTACA 778

Qy   1551 ATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAAC 1588
          ||||||||  ||   ||| ||||||| ||  ||   ||
Db    779 ATCTGTGTCCATTACAAAATTGAACCGTTGTTACACAC 816
```

RESULT 14
BI220592
LOCUS BI220592 681 bp mRNA linear EST 11-JUL-2001
DEFINITION 602936079F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5099346 5', mRNA sequence.
ACCESSION BI220592
VERSION BI220592.1 GI:14674036
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 681)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11238 row: n column: 19
High quality sequence stop: 681.

FEATURES Location/Qualifiers
source 1. .681
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5099346"
/clone_lib="NCI_CGAP_Li9"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 167 a 165 c 193 g 156 t
ORIGIN

Query Match 32.5%; Score 577; DB 13; Length 681;
Best Local Similarity 92.2%; Pred. No. 1.6e-124;
Matches 630; Conservative 0; Mismatches 50; Indels 3; Gaps 2;

```
Qy    441 GTACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCTG 500
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      1 GCACGGTGTGGGAGCTGAGCGAGGAGGGCGTGTACGGCCAGGACTCGCCGCTGGAGCCCG 60

Qy    501 TGGCTGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTTAACGCCTGTAACCCGCACA 560
          |  | ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     61 TCTCCGGGGTCCTGGTACCGCCCGACGGGCCCGGGGCGCTCAACGCCTGTAACCCGCACA 120

Qy    561 CGAATTTCACGGTGCCCACGGTTTGGGGAAGCACCGTGCAAGTCTCTTGGTTGGCCCTCA 620
          | |||||||||||||||||||||||||||| ||||| |||||||| ||||||||||||||||
Db    121 CCAATTTCACGGTGCCCACGGTTTGGGGGAGCACGGTGCAAGTATCTTGGTTGGCCCTCA 180

Qy    621 TCCAACGCGGCGGGGGCTGCACCTTCGCAGACAAGATCCATCTGGCTTATGAGAGATGGG 680
          |||||||||| || |||||||||||||| ||||||||||||||||||  |||||| |||
Db    181 TCCAACGCGGTGGAGGCTGCACCTTCGCGGACAAGATCCATCTGGCTTCAGAGAGAGGGG 240
```

```
Qy     681 CGTCTGGAGCCGTCATCTTTAACTTCCCCGGGACCCGCAATGAGGTCATCCCCATGTCTC 740
           | |||||||| |||||||||||||||||| |||||||||||||||||||||||||||||||
Db     241 CTTCTGGAGCGGTCATCTTTAACTTCCCTGGGACCCGCAATGAGGTCATCCCCATGTCTC 300

Qy     741 ACCCGGGTGCAGTAGACATTGTTGCAATCATGATCGGCAATCTGAAAGGCACAAAAATTC 800
           |||||||||| |  |||||||||||||||||||| |||||||||||||| ||||||||||
Db     301 ACCCGGGTGCTGGGGACATTGTTGCAATCATGATTGGCAATCTGAAAGGAACAAAAATTC 360

Qy     801 TGCAATCTATTCAAAGAGGCATACAAGTGACAATGGTCATAGAAGTAGGGAAAAAACATG 860
           ||||  |||||||||||||||| ||||| |||||||||| |||||||||  ||||||||
Db     361 TGCAGTCTATTCAAAGAGGCATCCAAGTCACAATGGTCATCGAAGTAGGG--AAAACATG 418

Qy     861 GCCCTTGGGTG-AATCACTATTCAATTTTTTCGTTTCTGTGTCCTTTTTTATTATTACG 919
           |||||||||||| ||||| |||| |||||| |||||||||||||||||| || ||||||
Db     419 GCCCTTGGGTGAAATCATTATTCAATTTTCTTCGTTTCTGTGTCCTTTTTCATAATTACG 478

Qy     920 GCGGCAACTGTGGGCTATTTTATCTTTTATTCTGCTCGAAGGCTACGGAATGCAAGAGCT 979
           || ||||| |||||||||| ||||||||||||||||||  |||| ||||||||||||
Db     479 GCAGCAACCGTGGGCTATTTCATCTTTTATTCTGCTCGAAGATTACGAAATGCAAGAGCT 538

Qy     980 CAAAGCAGGAAGCAGAGGCAATTAAAGGCAGATGCTAAAAAAGCTATTGGAAGGCTTCAA 1039
           |||||||||||||||||| |||||||||||||||||||||||||||||| ||||||
Db     539 CAAAGCAGGAAGCAGAGGCAGTTAAAGGCAGATGCTAAAAAAGCTATTGGAAAGCTTCAG 598

Qy    1040 CTACGCACACTGAAACAAGGAGACAAGGAAATTGGCCCTGATGGAGATAGTTGTGCTGTG 1099
           || |||||  |||||||||||||||||||||||| |||||||||||| |||||||||
Db     599 CTGCGCACCTTGAAACAAGGAGACAAGGAAATTGGGCCTGATGGAGATAGCTGTGCTGTG 658

Qy    1100 TGCATTGAATTGTATAAACCAAA 1122
           ||||||||  | ||||| |||||
Db     659 TGCATTGAGCTCTATAAGCCAAA 681
```

```
RESULT 15
AV692456
LOCUS       AV692456                683 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV692456 GKC Homo sapiens cDNA clone GKCCBD09 5', mRNA sequence.
ACCESSION   AV692456
VERSION     AV692456.1  GI:10294319
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GKCCBD09"
                     /clone_lib="GKC"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
```

```
BASE COUNT      225 a    127 c    126 g    200 t      5 others
ORIGIN

  Query Match           31.8%;  Score 564.4;  DB 10;  Length 683;
  Best Local Similarity  98.0%;  Pred. No. 1.4e-121;
  Matches  591;  Conservative    0;  Mismatches    9;  Indels    3;  Gaps    2;

Qy    1172 GTTGACCCATGGCTGTTAAAACACAAGACTTGCCCCATGTGCAAATGTGACATACTCAAA 1231
           |||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db       1 GTTGACCCATGGCTGTTAGAACACAGGACTTGCCCCATGTGCAAATGTGACATACTCAAA 60

Qy    1232 GCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCC 1291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTTTGGGAATTGAGGTGGATGTTGAAGATGGATCAGTGTCTTTACAAGTCCCTGTATCC 120

Qy    1292 AATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCA 1351
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AATGAAATATCTAATAGTGCCTCCTCCCATGAAGAGGATAATCGCAGCGAGACCGCATCA 180

Qy    1352 TCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCA 1411
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCTGGATATGCTTCAGTACAGGGAACAGATGAACCGCCTCTGGAGGAACACGTGCAGTCA 240

Qy    1412 ACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATT 1471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACAAATGAAAGTCTACAGCTGGTAAACCATGAAGCAAATTCTGTGGCAGTGGATGTTATT 300

Qy    1472 CCTCATGTNGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTT 1531
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTCATGTTGACAACCCAACCTTTGAAGAAGACGAAACTCCTAATCAAGAGACTGCTGTT 360

Qy    1532 CGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGA 1591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGAGAAATTAAATCTTAAAATCTGTGTAAATAGAAAACTTGAACCATTAGTAATAACAGA 420

Qy    1592 ACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATTTAATAAAATAAGAGTG 1651
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     421 ACTGCCAATCAGGGCCTAGTTTCTATTAATAAATTGGATAAATNTAATAAAATAAGAGTG 480

Qy    1652 ATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAAAATGGCTTAAAAATATTT 1711
           |||||||||||||||||||||||||||||||||||||||||||  |  | |||||||||
Db     481 ATACTGAAAGTGCTCAGATGACTAATATTATGCTATAGTTAAA--TGGCTTAAAATATTT 538

Qy    1712 AACCTGTTAACTTTTTTCCACCAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCT 1771
           |||||||||||| |||||| |||||||||||||||||||||||||||||||||||||||||
Db     539 AACCTGTTAACTNTTTTCCA-CAAACTCATTATAATATTTTTCATAGGCAAGTTTCCTCT 597

Qy    1772 CAG 1774
           |||
Db     598 CAG 600
```

```
Search completed: April 26, 2003, 02:27:20
Job time : 1767 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 22, 2003, 08:27:01 ; Search time 42 Seconds
(without alignments)
816.561 Million cell updates/sec

Title: US-09-854-300-8
Perfect score: 2275
Sequence: 1 MGPPPGAGVSCRGGCGFSRL.....PTFEEDETPNQETAVREIKS 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1134	49.8	225	10	US-09-764-864-850	Sequence 850, App
2	838	36.8	165	10	US-09-764-864-1306	Sequence 1306, Ap
3	656.5	28.9	400	10	US-09-935-390A-20	Sequence 20, Appl
4	468.5	20.6	276	10	US-09-935-390A-35	Sequence 35, Appl
5	396	17.4	147	10	US-09-764-864-1272	Sequence 1272, Ap
6	246.5	10.8	381	9	US-09-798-889-65	Sequence 65, Appl
7	227.5	10.0	296	10	US-09-925-300-1231	Sequence 1231, Ap
8	223	9.8	259	9	US-09-798-889-127	Sequence 127, App
9	137.5	6.0	337	10	US-09-828-303-18	Sequence 18, Appl
10	130.5	5.7	199	10	US-09-764-864-1291	Sequence 1291, Ap
11	130.5	5.7	207	10	US-09-925-302-638	Sequence 638, App
12	130	5.7	461	10	US-09-764-864-1288	Sequence 1288, Ap
13	130	5.7	636	10	US-09-764-864-832	Sequence 832, App
14	129.5	5.7	201	10	US-09-764-864-836	Sequence 836, App
15	127.5	5.6	664	9	US-10-176-847-40	Sequence 40, Appl
16	127.5	5.6	664	10	US-09-898-533-2	Sequence 2, Appli
17	126.5	5.6	685	10	US-09-745-763-11	Sequence 11, Appl
18	123.5	5.4	291	10	US-09-764-864-1293	Sequence 1293, Ap
19	118.5	5.2	171	10	US-09-764-864-840	Sequence 840, App
20	118.5	5.2	171	10	US-09-764-864-1295	Sequence 1295, Ap
21	105.5	4.6	124	10	US-09-764-864-803	Sequence 803, App
22	105.5	4.6	124	10	US-09-764-864-1264	Sequence 1264, Ap
23	105	4.6	155	10	US-09-925-300-1159	Sequence 1159, Ap
24	105	4.6	205	10	US-09-764-864-1260	Sequence 1260, Ap
25	105	4.6	224	10	US-09-764-864-799	Sequence 799, App
26	103.5	4.5	662	10	US-09-799-777-53	Sequence 53, Appl
27	103.5	4.5	1863	9	US-09-734-672-2	Sequence 2, Appli
28	103.5	4.5	1863	9	US-09-734-672-4	Sequence 4, Appli
29	103.5	4.5	1863	9	US-09-734-672-6	Sequence 6, Appli
30	103.5	4.5	1863	9	US-09-982-828-2	Sequence 2, Appli
31	103.5	4.5	1863	9	US-09-982-828-4	Sequence 4, Appli
32	103.5	4.5	1863	9	US-09-982-828-6	Sequence 6, Appli
33	103.5	4.5	1863	9	US-10-022-819-2	Sequence 2, Appli
34	102.5	4.5	426	10	US-09-764-864-809	Sequence 809, App
35	102.5	4.5	664	10	US-09-780-525-2	Sequence 2, Appli
36	102	4.5	26926	9	US-09-759-508B-2	Sequence 2, Appli
37	101.5	4.5	825	9	US-10-121-988-161	Sequence 161, App
38	101.5	4.5	826	9	US-10-121-988-47	Sequence 47, Appl
39	101.5	4.5	826	10	US-09-894-998-47	Sequence 47, Appl
40	100	4.4	600	10	US-09-764-864-1282	Sequence 1282, Ap
41	100	4.4	615	10	US-09-925-301-1094	Sequence 1094, Ap
42	100	4.4	759	10	US-09-764-864-823	Sequence 823, App
43	98.5	4.3	439	10	US-09-764-864-849	Sequence 849, App
44	97	4.3	266	10	US-09-764-864-834	Sequence 834, App
45	96.5	4.2	811	10	US-09-840-787-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-09-764-864-850
; Sequence 850, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 850
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-850

Query Match 49.8%; Score 1134; DB 10; Length 225;
Best Local Similarity 99.1%; Pred. No. 5.9e-106;
Matches 218; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy    209 FFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDK 268
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 FFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDK 65

Qy    269 EIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIE 328
          ||||||||||||||||||||||||||||||||||||||||||||:|:|||||||||||||||
Db     66 EIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIE 125

Qy    329 VDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESL 388
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 VDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESL 185

Qy    389 QLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 428
          |||||||||||||||||||||||||||||||||||||||||
Db    186 QLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 225
```

RESULT 2
US-09-764-864-1306

```
; Sequence 1306, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1306
;   LENGTH: 165
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1306

  Query Match             36.8%;  Score 838;  DB 10;  Length 165;
  Best Local Similarity   97.5%;  Pred. No. 2.3e-76;
  Matches  157;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     268 KEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGI 327
           : |||||||||||||||||||||||||||||||||||||||||||:|:|||||||||||||
Db       5 RXIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGI  64

Qy     328 EVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNES 387
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65 EVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNES 124

Qy     388 LQLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 428
           |||||||||||||||||||||||||||||||||||||||||
Db     125 LQLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 165


RESULT 3
US-09-935-390A-20
; Sequence 20, Application US/09935390A
; Patent No. US20020076761A1
;   GENERAL INFORMATION:
;        APPLICANT: Escobedo, Jaime
;                   Quianjin, Hu
;                   Garcia, Pablo
;                   Williams, Lewis T.
;                   Kothakota, Srinivas
;        TITLE OF INVENTION: Secreted Human Proteins
;        NUMBER OF SEQUENCES: 38
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Chiron Corporation
;             STREET: 4560 Horton Street
;             CITY: Emeryville
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94608-2916
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/935,390A
;             FILING DATE: 22-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/988,671
;             FILING DATE: 1997-12-11
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jane E. R. Potter
;             REGISTRATION NUMBER: 33,332
;             REFERENCE/DOCKET NUMBER: 1369.002
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (510) 923-2718
;             TELEFAX: (510) 655-3542
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 20:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 400 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: No. US20020076761A1e
;        SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-935-390A-20

  Query Match             28.9%;  Score 656.5;  DB 10;  Length 400;
  Best Local Similarity   37.9%;  Pred. No. 1.8e-57;
  Matches  161;  Conservative   65;  Mismatches  148;  Indels   51;  Gaps   13;

Qy       7 AGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVW-TAYLNVSWRVPHTGVNRTVWEL 65
           |||  ||      :||   ||||:   ||:||    | :| :|: :  | |  | ||| :
Db       8 AGVGARG------VLALA-LLALALCVPGARGRALEWFSAVVNIEYVDPQT--NLTVWSV 58

Qy      66 SEEGVYGQDSPLEPVAGVLVPPDGPGA-LNACNPHTNFTVPTVWGSTVQVSWLALIQRGG 124
           || | :|  || |    |::  |  ||  |  | | | | ||   |       |:||: | |
Db      59 SESGRFGDSSPKEGAHGLVGVPWAPGGDLEGCAPDTRFFVPEP-GGRGAAPWVALVAR-G 116

Qy     125 GCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQ 184
           |||| ||: :|  | ||  |::|     |  :|||| |  :|| |||   || :||: :|
Db     117 GCTFKDKVLVAARRNASAVVLYNEERYGNITLPMSHAGTGNIVVIMISYPKGREILELVQ 176

Qy     185 RGIQVTMVIEVGKKH-GPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRK 243
           :|| ||| | || :|   :::  |: ||:::|  :   :: : |||  :|     :|
Db     177 KGIPVTMTIGVGTRHVQEFISGQSVVFVAIAFITMMIISLAWLIFYYIQRFLYTGSQIGS 236

Qy     244 QRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTC 303
           |   | : || ||:| | |:| |:| |  | ::|||||| :|  |::||| | ||||: |
Db-    237 QSHRK-ETKKVIGQLLLHTVKHGEKGIDVDAENCAVCIENFKVKDIIRILPCKHIFHRIC 295

Qy     304 VDPWLLKHKTCPMCKCDILKALGI---EVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           :||||| |:|||||| |::||||      ||::       |  :  :| : :  :|: |:
Db     296 IDPWLLDHRTCPMCKLDVIKALGYWGEPGDVQEMPAPESPPGRDPAANLSLALPDDDGSD 355

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           :|   ||      | |   |                        :|:|: |   |
Db     356 DSSPPSAS-------PAESEPQC-----------------------DPSFKGDAGEN-- 382

Qy     421 TAVRE 425
           ||: |
Db     383 TALLE 387


RESULT 4
US-09-935-390A-35
; Sequence 35, Application US/09935390A
; Patent No. US20020076761A1
;   GENERAL INFORMATION:
;        APPLICANT: Escobedo, Jaime
;                   Quianjin, Hu
;                   Garcia, Pablo
;                   Williams, Lewis T.
;                   Kothakota, Srinivas
;        TITLE OF INVENTION: Secreted Human Proteins
;        NUMBER OF SEQUENCES: 38
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Chiron Corporation
;             STREET: 4560 Horton Street
;             CITY: Emeryville
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94608-2916
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
```

```
;              OPERATING SYSTEM: DOS
;              SOFTWARE: FastSEQ for Windows Version 2.0
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/935,390A
;              FILING DATE: 22-Aug-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/988,671
;              FILING DATE: 1997-12-11
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Jane E. R. Potter
;              REGISTRATION NUMBER: 33,332
;              REFERENCE/DOCKET NUMBER: 1369.002
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (510) 923-2718
;              TELEFAX: (510) 655-3542
;              TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 35:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 276 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: No. US20020076761A1e
;         SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-935-390A-35

  Query Match             20.6%;  Score 468.5;  DB 10;  Length 276;
  Best Local Similarity   52.6%;  Pred. No. 8.5e-39;
  Matches   90;  Conservative   28;  Mismatches   52;  Indels    1;  Gaps    1;

Qy    158 MSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPW-VNHYSIFFVSVSFF 216
          |:|||  ||:|:||  |:|  ||  ::: | | | | || :  |   :  |: |||:||
Db      1 MTHPGTGDIIAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGSLVFVSISFI 60

Qy    217 IITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDS 276
          ::   :  : |||  :::|   |: | ||:|   ||||| :|  ||:|:|||| || |
Db     61 VLMIISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKETDPDFDH 120

Qy    277 CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGI 327
          |||||| || ||:|||| | |:|||:|||||| :| ||||||  :|||||||
Db    121 CAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNILKALGI 171


RESULT 5
US-09-764-864-1272
; Sequence 1272, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1272
;   LENGTH: 147
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-1272

  Query Match             17.4%;  Score 396;  DB 10;  Length 147;
  Best Local Similarity   56.9%;  Pred. No. 6.3e-32;
  Matches   70;  Conservative   24;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    202 WVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLR 261
          :|:  |: |||:|| ::   :: : :||  :| | | |: | ||:|   ||||| :||:|
Db     10 YVSRTSVVFVSISFIVLMIISLAWLVFYYIQRFRYANARDRNQRRLGDAAKKAISKLQIR 69

Qy    262 TLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDI 321
          |:|:|||   | |:|||||| |||||:|||| | |:|||:|||||| |:|||||  :|
Db     70 TIKKGDKETESDFDNCAVCIEGYKPNDVVRILPCRHLFHKSCVDPWLLDHRTCPMCKMNI 129

Qy    322 LKA 324
          ||:
Db    130 LKS 132


RESULT 6
US-09-798-889-65
; Sequence 65, Application US/09798889
; Publication No. US20030004324A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 31 Human secreted proteins
;  FILE REFERENCE: PZ026P1
;  CURRENT APPLICATION NUMBER: US/09/798,889
;  CURRENT FILING DATE:  2001-03-06
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-09
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,687
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,696
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  NUMBER OF SEQ ID NOS: 185
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
;   LENGTH: 381
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-798-889-65

  Query Match             10.8%;  Score 246.5;  DB 9;  Length 381;
  Best Local Similarity   24.1%;  Pred. No. 3.2e-16;
  Matches   83;  Conservative   58;  Mismatches  132;  Indels   71;  Gaps   14;

Qy     71 YGQDSPLEPVAGVLV--PPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTF 128
          :|   | | : | |:   |:     ||| |     ||          ::: ||:|   | |
Db     55 FGYRLPAEGLKGFLINSKPE-----NACEP----IVPPPVKDNSSGTFIVLIRR-LDCNF 104

Qy    129 ADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVDIVAIMIGNLKGTKILQ-- 181
            |:  |       |:: |     :::| |           :|| :: ||      :
Db    105 DIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEVLKKIDIPSVFIGESSANSLKDEF 162

Qy    182 SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQS 241
          : ::|  : :| |             : :  : | ||    :   : :   :    | ::
Db    163 TYEKGGHLILVPEFS---------LPLEYYLIPFLIIVGICLILIVIFMITKFVQDRHRA 213

Qy    242 RKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHK 301
          |: |  |   ||    | :   |:||     : | ||:|:: |:  | :||| |:| :|
Db    214 RRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDEYEDGDKLRILPCSHAYHC 264

Qy    302 TCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVSLQVPV---------- 342
          |||||| |  ||||:||  :: : |       :   |:  |:   |:
Db    265 KCVDPWLTKTKKTCPVCKQKVVPSQGDSDSDTDSSQEENEVTEHTPLLRPLASVSAQSFG 324

Qy    343 ------SNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEH 380
                |::    :| :|||:  :| ||          |  : ||
Db    325 ALSESRSHQNMTESSDYEEDDNEDTDSS-------DAENEINEH 361


RESULT 7
US-09-925-300-1231
; Sequence 1231, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
```

```
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1231
;   LENGTH: 296
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1231

  Query Match             10.0%;  Score 227.5;  DB 10;  Length 296;
  Best Local Similarity   23.2%;  Pred. No. 1.8e-14;
  Matches   70;  Conservative   54;  Mismatches  119;  Indels   59;  Gaps   10;

Qy     111 TVQVSWLALIQRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVD 165
           |: :    :: |   | |  |:  |        |:: |     :::| |           :|
Db       2 TIHLXTFIVLIRRLDCNFDIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEVLKKID 59

Qy     166 IVAIMIGNLKGTKILQ--SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATV 223
           | :: ||      :    : ::|  : :| |             : :  : | ||    :
Db      60 IPSVFIGESSANSLKDEFTYEKGGHLILVPEFS---------LPLEYYLIPFLIIVGICL 110

Qy     224 GYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIEL 283
           : :   :    | ::|: |  |   ||    | :   |:||     : | ||:|::
Db     111 ILIVIFMITKFVQDRHRARRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDE 161

Qy     284 YKPNDLVRILTCNHIFHKTCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVS 337
           |:  | :||| |:| :|  |||||| |  ||||:||  :: : |       :   |:  |:
Db     162 YEDGDKLRILPCSHAYHCKCVDPWLTKTKKTCPVCKQKVVPSQGDSDSDTDSSQEENEVT 221

Qy     338 LQVPV------------------SNEISNSASSHEEDNRSETASSGYASVQGTDEPPLE 378
             |:                   |::    :| :|||:  :| ||         |  :
Db     222 EHTPLLRPLASVSAQSFGALSESRSHQNMTESSDYEEDDNEDTDSS-------DAENEIN 274

Qy     379 EH 380
           ||
Db     275 EH 276


RESULT 8
US-09-798-889-127
; Sequence 127, Application US/09798889
; Publication No. US20030004324A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 31 Human secreted proteins
;  FILE REFERENCE: PZ026P1
;  CURRENT APPLICATION NUMBER: US/09/798,889
;  CURRENT FILING DATE:  2001-03-06
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-09
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,687
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,696
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
;  NUMBER OF SEQ ID NOS: 185
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 127
;   LENGTH: 259
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-798-889-127

  Query Match             9.8%;  Score 223;  DB 9;  Length 259;
  Best Local Similarity   24.0%;  Pred. No. 4e-14;
  Matches   65;  Conservative   49;  Mismatches  105;  Indels   52;  Gaps    9;

Qy     126 CTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVDIVAIMIGNLKGTKIL 180
           | |  |:  |       |:: |     :::| |           :|| :: ||      :
Db       6 CNFDIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEVLKKIDIPSVFIGESSANSLK 63

Qy     181 Q--SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNAR 238
           : ::|  : :| |             : :  : | ||    :   : :   :    |
Db      64 DEFTYEKGGHLILVPEFS---------LPLEYYLIPFLIIVGICLILIVIFMITKFVQDR 114

Qy     239 AQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHI 298
           ::|: |  |   ||    | :   |:||     : | ||:|:: |:  | :||| |:|
Db     115 HRARRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDEYEDGDKLRILPCSHA 165

Qy     299 FHKTCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVSLQVPV---------- 342
           :| |||||| |  ||||:||  :: : |       :   |:  |:   |:
Db     166 YHCKCVDPWLTKTKKTCPVCKQKVVPSQGDSDSDTDSSQEENEVTEHTPLLRPLASVSAQ 225

Qy     343 ---------SNEISNSASSHEEDNRSETASS 364
                   |::    :| :|||:  :| ||
Db     226 SFGALSESRSHQNMTESSDYEEDDNEDTDSS 256


RESULT 9
US-09-828-303-18
; Sequence 18, Application US/09828303
; Patent No. US20020102695A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND
;  TITLE OF INVENTION:  METHODS OF USE IN PLANTS
;  FILE REFERENCE: 16313-0030
;  CURRENT APPLICATION NUMBER: US/09/828,303
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Physcomitrella patens
US-09-828-303-18

  Query Match             6.0%;  Score 137.5;  DB 10;  Length 337;
  Best Local Similarity   24.4%;  Pred. No. 2.5e-05;
  Matches   39;  Conservative   26;  Mismatches   78;  Indels   17;  Gaps    5;

Qy     273 DGD------SCAVCIELYKPNDLVRIL-TCNHIFHKTCVDPWLLKHKTCPMCKCDILKAL 325
           ||:        ||||:| ::  :  | |  |:| ||  |:| ||  | |||:|:  :  |
Db      78 DGERAKRVFECAVCLEEFELGEKGRTLPKCDHSFHLDCIDMWLHSHSTCPLCRTSV-GAD 136

Qy     326 GIEVDVEDGSV-------SLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLE 378
             |   | :|         :: ||  ::      :    :|    | |   :   :  |
Db     137 ETEKKTEAATVMQISEPPQMEAPVMGDVGAPFMAAMRASRRSQRSRG--QLPALNSSPRG 194

Qy     379 EHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPN 418
           :  | |      || :  :   :         :: |||:
Db     195 NSLPRTAEDQGGENHRRSGTSETAVAVDQQQNIKDYETPS 234


RESULT 10
US-09-764-864-1291
; Sequence 1291, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1291
;   LENGTH: 199
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (187)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1291

  Query Match             5.7%;  Score 130.5;  DB 10;  Length 199;
  Best Local Similarity   33.8%;  Pred. No. 5.5e-05;
  Matches   24;  Conservative   16;  Mismatches   26;  Indels    5;  Gaps    1;

Qy    252 KKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKH 311
          |: |  | :|:  : |        :|:|||  |   : :| | |:| :|  |:| || ::
Db    125 KEQIDNLAMRSPGENDAL-----KTCSVCITEYTEGNKLRKLPCSHEYHVHCIDRWLSEN 179

Qy    312 KTCPMCKCDIL 322
          |||:|:  :|
Db    180 STCPICRXAVL 190


RESULT 11
US-09-925-302-638
; Sequence 638, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 638
;   LENGTH: 207
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-302-638

  Query Match             5.7%;  Score 130.5;  DB 10;  Length 207;
  Best Local Similarity   33.0%;  Pred. No. 5.9e-05;
  Matches   29;  Conservative   16;  Mismatches   32;  Indels   11;  Gaps    3;

Qy    239 AQSRKQRQLKADAK----KAIGRLQLRTLKQGDKEIGPDGD-----SCAVCIELYKPNDL 289
          :|:|   ||  : :    : || :|   | :|  : | ||        | :|:  :   |
Db    107 SQTRLATQLTEEEQIRIAQRIGLIQ--HLPKGVYDPGRDGSEKKIRECVICMMDFVYGDP 164

Qy    290 VRILTCNHIFHKTCVDPWLLKHKTCPMC 317
          :| | | ||:|  |:| ||::  ||| |
Db    165 IRFLPCMHIYHLDCIDDWLMRSFTCPSC 192


RESULT 12
US-09-764-864-1288
; Sequence 1288, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1288
;   LENGTH: 461
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (83)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (250)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1288

  Query Match             5.7%;  Score 130;  DB 10;  Length 461;
  Best Local Similarity   28.7%;  Pred. No. 0.00023;
  Matches   33;  Conservative   17;  Mismatches   49;  Indels   16;  Gaps    3;

Qy    220 AATVGY----FIFYSARRLRNARAQSRKQRQLKADA----------KKAIGRLQLRTLK 264
          |  : |    |: | |   | |:|        |:: |          |::|  |    :
Db    342 AEAISYVDPQFLTYMALEERLAQAMETALAHLESLAVDVEVANPPASKESIDALPEILVT 401

Qy    265 QGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKC 319
          :    :| :   | :|   |   ::   | |:| ||| ||  || |  |||:|:|
Db    402 EDHGAVGQE-MCCPICCSEYVKGEVATELPCHHYFHKPCVSIWLQKSGTCPVCRC 455


RESULT 13
US-09-764-864-832
; Sequence 832, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 832
;   LENGTH: 636
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-832

  Query Match             5.7%;  Score 130;  DB 10;  Length 636;
  Best Local Similarity   28.7%;  Pred. No. 0.00039;
  Matches   33;  Conservative   17;  Mismatches   49;  Indels   16;  Gaps    3;

Qy    220 AATVGY----FIFYSARRLRNARAQSRKQRQLKADA----------KKAIGRLQLRTLK 264
          |  : |    |: | |   | |:|        |:: |          |::|  |    :
Db    517 AEAISYVDPQFLTYMALEERLAQAMETALAHLESLAVDVEVANPPASKESIDALPEILVT 576

Qy    265 QGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKC 319
          :    :| :   | :|   |   ::   | |:| ||| ||  || |  |||:|:|
Db    577 EDHGAVGQE-MCCPICCSEYVKGEVATELPCHHYFHKPCVSIWLQKSGTCPVCRC 630


RESULT 14
US-09-764-864-836
; Sequence 836, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
```

```
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 836
;   LENGTH: 201
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-836

  Query Match             5.7%;  Score 129.5;  DB 10;  Length 201;
  Best Local Similarity   33.8%;  Pred. No. 7e-05;
  Matches   24;  Conservative   16;  Mismatches   26;  Indels    5;  Gaps    1;

Qy     252 KKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKH 311
           |: |  | :|:  :  |         :|:|||  |   : :| | |:| :|  |:| || ::
Db     127 KEQIDNLAMRSFGENDAL-----KTCSVCITEYTEGNKLRKLPCSHEYHVHCIDRWLSEN 181

Qy     312 KTCPMCKCDIL 322
           |||:|:  :|
Db     182 STCPICRRAVL 192


RESULT 15
US-10-176-847-40
; Sequence 40, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
;  APPLICANT: Veiby, Petter Ole
;  TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
;  TITLE OF INVENTION:  AND OVARIAN CANCER
;  FILE REFERENCE: MRI-039
;  CURRENT APPLICATION NUMBER: US/10/176,847
;  CURRENT FILING DATE:  2002-06-21
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40
;   LENGTH: 664
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-847-40

  Query Match             5.6%;  Score 127.5;  DB 9;  Length 664;
  Best Local Similarity   25.2%;  Pred. No. 0.00074;
  Matches   38;  Conservative   18;  Mismatches   44;  Indels   51;  Gaps    6;

Qy     275 DSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDG 334
           | ||:|   :  :   ||  ||| ||  |:  ||    |||||     : : || |::|
Db     545 DVCAICYHEPTTS--ARITPCNHYFHALCLRKWLYIQDTCPMCH----QKVYIEDDIKDN 598

Qy     335 SVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVNHE 394
           |          :||               ::|:         || |   ::   |:    : |
Db     599 S---------NVSN-------------NNGFI--------PPNETPEEAVREAAAESDRE 628

Qy     395 ANSVAVDVIPHVDNPTFEEDETPNQETAVRE 425
           |                |:| |   :   ||
Db     629 LN---------------EDDSTDCDDDVQRE 644


Search completed: April 22, 2003, 08:34:47
Job time : 43 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 22, 2003, 08:24:36 ; Search time 17 Seconds
(without alignments)
740.765 Million cell updates/sec

Title: US-09-854-300-8
Perfect score: 2275
Sequence: 1 MGPPPGAGVSCRGGCGFSRL.....PTFEEDETPNQETAVREIKS 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	468.5	20.6	276	2	US-08-786-606-5	Sequence 5, Appli
2	331	14.5	284	2	US-08-786-606-9	Sequence 9, Appli
3	253	11.1	381	2	US-08-867-057-3	Sequence 3, Appli
4	253	11.1	381	2	US-09-128-369-3	Sequence 3, Appli
5	246.5	10.8	381	2	US-08-867-057-1	Sequence 1, Appli
6	246.5	10.8	381	2	US-09-128-369-1	Sequence 1, Appli
7	152.5	6.7	50	4	US-09-052-089A-15	Sequence 15, Appl
8	144.5	6.4	40	4	US-09-046-894-44	Sequence 44, Appl
9	131.5	5.8	104	4	US-09-325-932A-49	Sequence 49, Appl
10	130	5.7	395	2	US-08-841-349-9	Sequence 9, Appli
11	127.5	5.6	664	4	US-09-268-140-2	Sequence 2, Appli
12	114.5	5.0	199	4	US-09-325-932A-46	Sequence 46, Appl
13	108.5	4.8	551	3	US-08-699-103B-25	Sequence 25, Appl
14	108.5	4.8	551	4	US-09-229-059-25	Sequence 25, Appl
15	107.5	4.7	469	2	US-08-968-751-2	Sequence 2, Appli
16	106.5	4.7	470	4	US-09-052-089A-2	Sequence 2, Appli
17	103.5	4.5	312	1	US-08-425-061-18	Sequence 18, Appl
18	103.5	4.5	312	2	US-08-825-886-18	Sequence 18, Appl
19	103.5	4.5	765	1	US-08-425-061-19	Sequence 19, Appl
20	103.5	4.5	765	2	US-08-825-886-19	Sequence 19, Appl
21	103.5	4.5	900	1	US-08-425-061-20	Sequence 20, Appl
22	103.5	4.5	900	2	US-08-825-886-20	Sequence 20, Appl
23	103.5	4.5	914	1	US-08-425-061-21	Sequence 21, Appl
24	103.5	4.5	914	2	US-08-825-886-21	Sequence 21, Appl
25	103.5	4.5	1202	1	US-08-425-061-22	Sequence 22, Appl
26	103.5	4.5	1202	2	US-08-825-886-22	Sequence 22, Appl
27	103.5	4.5	1363	1	US-08-425-061-23	Sequence 23, Appl
28	103.5	4.5	1363	2	US-08-825-886-23	Sequence 23, Appl
29	103.5	4.5	1852	1	US-08-425-061-24	Sequence 24, Appl
30	103.5	4.5	1852	2	US-08-825-886-24	Sequence 24, Appl
31	103.5	4.5	1863	1	US-08-598-591-2	Sequence 2, Appli
32	103.5	4.5	1863	1	US-08-480-784-2	Sequence 2, Appli
33	103.5	4.5	1863	1	US-08-483-553-2	Sequence 2, Appli
34	103.5	4.5	1863	1	US-08-487-002-2	Sequence 2, Appli
35	103.5	4.5	1863	1	US-08-483-554B-2	Sequence 2, Appli
36	103.5	4.5	1863	1	US-08-798-691-2	Sequence 2, Appli
37	103.5	4.5	1863	1	US-08-798-691-4	Sequence 4, Appli
38	103.5	4.5	1863	1	US-08-798-691-6	Sequence 6, Appli
39	103.5	4.5	1863	1	US-08-488-011B-2	Sequence 2, Appli
40	103.5	4.5	1863	2	US-08-603-753D-2	Sequence 2, Appli
41	103.5	4.5	1863	3	US-08-825-487A-2	Sequence 2, Appli
42	103.5	4.5	1863	3	US-08-825-487A-4	Sequence 4, Appli
43	103.5	4.5	1863	3	US-08-825-487A-6	Sequence 6, Appli
44	103.5	4.5	1863	4	US-09-074-476-2	Sequence 2, Appli
45	103.5	4.5	1863	4	US-09-074-476-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-786-606-5
; Sequence 5, Application US/08786606
; Patent No. 5861495
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Au-Young, Janice
; APPLICANT: Coleman, Roger
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,606
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0173 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 276 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-786-606-5

Query Match 20.6%; Score 468.5; DB 2; Length 276;
Best Local Similarity 52.6%; Pred. No. 9e-44;
Matches 90; Conservative 28; Mismatches 52; Indels 1; Gaps 1;

```
Qy     158 MSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPW-VNHYSIFFVSVSFF 216
           |:|||  ||:|:||  |:|  ||  ::: | | | | || :  |   :  |: |||:||
Db       1 MTHPGTGDIIAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGSLVFVSISFI 60

Qy     217 IITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDS 276
           ::   :  : |||  :::|   |: | ||:|   ||||| :|  ||:|:||||  || |
Db      61 VLMIISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKETDPDFDH 120

Qy     277 CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGI 327
           |||||| || ||:|||| | |:|||:|||||| :| |||||| :|||||||
Db     121 CAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNILKALGI 171
```

RESULT 2
US-08-786-606-9
; Sequence 9, Application US/08786606
; Patent No. 5861495
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Au-Young, Janice
; APPLICANT: Coleman, Roger
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,606
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0173 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 157535
US-08-786-606-9

Query Match 14.5%; Score 331; DB 2; Length 284;
Best Local Similarity 34.8%; Pred. No. 2.3e-28;
Matches 80; Conservative 39; Mismatches 77; Indels 34; Gaps 5;

```
Qy     165 DIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWV---NHYSIFFVSVSFFIITAA 221
           :| |::    |  :  :: :|  ||: |  |::    :   |  |: |||:||  |
Db      14 NIAAVITYQNIGQDLSLTLDKGYNVTISIIEGRRGVRTISSLNRTSVLFVSISF--IVDD 71

Qy     222 TVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCI 281
           : : |||  :| |  :|: :: | | :  |||| ::  :| |  | |   | | ||:||
Db      72 ILCWLIFYYIQRFRYMQAKDQQSRNLCSVTKKAIMKIPTKTGKFSD-EKDLDSDCCAICI 130

Qy     282 ELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALG-------------- 326
           | ||| | :||| | | ||| |:||||::|:|||||| |:||  |
Db     131 EAYKPTDTIRILPCKHEFHKNCIDPWLIEHRTCPMCKLDVLKFYGYVVGDQIYQTPSPQH 190

Qy     327 ------------IEVDVEDGSVSLQVPVSNEISNSASSHE-EDNRSETAS 363
                       | | | |   ||   :: :|: | ||  : :|| ::|
Db     191 TAPIASIEEVPVIVVAVPHGPQPLQPLQASNMSSFAPSHYFQSSRSPSSS 240
```

RESULT 3
US-08-867-057-3
; Sequence 3, Application US/08867057
; Patent No. 5840535
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW ZINC RING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,057
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0311 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1321818
US-08-867-057-3

Query Match 11.1%; Score 253; DB 2; Length 381;
Best Local Similarity 24.0%; Pred. No. 2.1e-19;
Matches 84; Conservative 60; Mismatches 140; Indels 66; Gaps 14;

```
Qy      71 YGQDSPLEPVAGVLV--PPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTF 128
           :|   | | : | |:   |:     |||| |      |          ::: ||:|   | |
Db      55 FGYRLPAEGLKGPLINSKPE-----NACEP----IAPPPLRDNSSTAFIVLIRR-LECNF 104

Qy     129 ADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVDIVAIMIGNLKGTKILQ-- 181
             |:  |       |:: |     :::| |           :|| :: ||      : :
Db     105 DIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEILKKIDIPSVFIGEASANSLKEEF 162

Qy     182 SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQS 241
           : ::|  | :: |          : :  : | ||    :   : :  :   | ::
Db     163 TYEKGGHVVLIPEFS---------LPLEYYLIPFLIIVGICLILIVIFMITKFVQDRHRA 213
```

```
Qy     242 RKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHK 301
Db     214 RRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDEYEDGDKLRILPCSHAYHC 264

Qy     302 TCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVSLQVPV------------- 342
Db     265 KCVDPWLTKTKKTCPVCKQKVVPSQGDSDSETDSSQEENEVSENTPLLRPLASVSTQSFG 324

Qy     343 ------SNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNE 386
Db     325 ALSESHSHQNMTESSEYEEDDNDNIDSSDAES--GVNEESVVVQLQPNDE 372


RESULT 4
US-09-128-369-3
; Sequence 3, Application US/09128369
; Patent No. 5998372
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NEW ZINC RING PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/128,369
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/867,057
;      FILING DATE:  06/02/1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0311 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1321818
US-09-128-369-3

  Query Match             11.1%;  Score 253;  DB 2;  Length 381;
  Best Local Similarity   24.0%;  Pred. No. 2.1e-19;
  Matches   84;  Conservative   60;  Mismatches  140;  Indels   66;  Gaps   14;

Qy      71 YGQDSPLEPVAGVLV--PPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTF 128
Db      55 FGYRLPAEGLKGFLINSKPE-----NACEP----IAPPPLRDNSSTAFIVLIRR-LECNF 104

Qy     129 ADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVDIVAIMIGNLKGTKILQ-- 181
Db     105 DIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEILKKIDIPSVFIGEASANSLKEEF 162

Qy     182 SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQS 241
```

```
Db     163 TYEKGGHVVLIPEFS---------LPLEYYLIPFLIIVGICLILIVIFMITKFVQDRHRA 213

Qy     242 RKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHK 301
Db     214 RRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDEYEDGDKLRILPCSHAYHC 264

Qy     302 TCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVSLQVPV------------- 342
Db     265 KCVDPWLTKTKKTCPVCKQKVVPSQGDSDSETDSSQEENEVSENTPLLRPLASVSTQSFG 324

Qy     343 ------SNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNE 386
Db     325 ALSESHSHQNMTESSEYEEDDNDNIDSSDAES--GVNEESVVVQLQPNDE 372


RESULT 5
US-08-867-057-1
; Sequence 1, Application US/08867057
; Patent No. 5840535
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NEW ZINC RING PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,057
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0311 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BMARNOT02
;      CLONE:  104119
US-08-867-057-1

  Query Match             10.8%;  Score 246.5;  DB 2;  Length 381;
  Best Local Similarity   24.1%;  Pred. No. 1.1e-18;
  Matches   83;  Conservative   58;  Mismatches  132;  Indels   71;  Gaps   14;

Qy      71 YGQDSPLEPVAGVLV--PPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTF 128
Db      55 FGYRLPAEGLKGFLINSKPE-----NACEP----IVPPPVKDNSSGTFIVLIRR-LDCNF 104

Qy     129 ADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVDIVAIMIGNLKGTKILQ-- 181
Db     105 DIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEVLKKIDIPSVFIGESSANSLKDEF 162
```

```
Qy     182 SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQS 241
           : ::|  : :| |            : :  : | ||     :  :   :    | ::
Db     163 TYEKGGHLILVPEFS---------LPLEYYLIPFLIIVGICLILIVIFMITKFVQDRHRA 213

Qy     242 RKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHK 301
           |: |  |   ||    | :   |:||     : | ||:|:: |:  | :||| |:| :|
Db     214 RRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDEYEDGDKLRILPCSHAYHC 264

Qy     302 TCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVSLQVPV------------ 342
           |||||| |  ||||:||  :: : |      :   |:  |:   |:
Db     265 KCVDPWLTKTKKTCPVCKQKVVPSQGDSDSDTDSSQEENEVTEHTPLLRPLASVSAQSFG 324

Qy     343 ------SNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEH 380
                 |::    :| :|||:  :| ||         |  : ||
Db     325 ALSESRSHQNMTESSDYEEDDNEDTDSS-------DAENEINEH 361


RESULT 6
US-09-128-369-1
; Sequence 1, Application US/09128369
; Patent No. 5998372
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NEW ZINC RING PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/128,369
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/867,057
;      FILING DATE:  06/02/1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0311 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BMARNOT02
;      CLONE:  104119
US-09-128-369-1

  Query Match             10.8%;  Score 246.5;  DB 2;  Length 381;
  Best Local Similarity   24.1%;  Pred. No. 1.1e-18;
  Matches   83;  Conservative   58;  Mismatches  132;  Indels   71;  Gaps   14;

Qy      71 YGQDSPLEPVAGVLV--PPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTF 128
           :|   | | : | |:   |:     ||| |     ||          ::: ||:|   | |
Db      55 FGYRLPAEGLKGFLINSKPE-----NACEP----IVPPPVKDNSSGTFIVLIRR-LDCNF 104

Qy     129 ADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPG-----AVDIVAIMIGNLKGTKILQ-- 181
```

```
           |:  |       |:: |     :::| |          :|| :: ||      :
Db     105 DIKVLNAQRAGYKAAIVHNVDS--DDLISMGSNDIEVLKKIDIPSVFIGESSANSLKDEF 162

Qy     182 SIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQS 241
           : ::|  : :| |            : :  : | ||     :  :   :    | ::
Db     163 TYEKGGHLILVPEFS---------LPLEYYLIPFLIIVGICLILIVIFMITKFVQDRHRA 213

Qy     242 RKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHK 301
           |: |  |   ||    | :   |:||     : | ||:|:: |:  | :||| |:| :|
Db     214 RRNRLRKDQLKK----LPVHKFKKGD-----EYDVCAICLDEYEDGDKLRILPCSHAYHC 264

Qy     302 TCVDPWLLK-HKTCPMCKCDILKALG-----IEVDVEDGSVSLQVPV------------ 342
           |||||| |  ||||:||  :: : |      :   |:  |:   |:
Db     265 KCVDPWLTKTKKTCPVCKQKVVPSQGDSDSDTDSSQEENEVTEHTPLLRPLASVSAQSFG 324

Qy     343 ------SNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEH 380
                 |::    :| :|||:  :| ||         |  : ||
Db     325 ALSESRSHQNMTESSDYEEDDNEDTDSS-------DAENEINEH 361


RESULT 7
US-09-052-089A-15
; Sequence 15, Application US/09052089A
; Patent No. 6346605
;  GENERAL INFORMATION:
;       APPLICANT: Lee, Soo Y.
;                  Choi, Yongwon
;       TITLE OF INVENTION: SIGNAL TRANSDUCER FOR THE TNF RECEPTOR SUPER
;                           FAMILY, AND USES THEREOF
;       NUMBER OF SEQUENCES: 16
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: David A. Jackson, Esq.
;            STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                    Floor
;            CITY: Hackensack
;            STATE: New Jersey
;            COUNTRY: USA
;            ZIP: 07601
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/052,089A
;            FILING DATE: 31-Mar-1998
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Jackson Esq., David A.
;            REGISTRATION NUMBER: 26,742
;            REFERENCE/DOCKET NUMBER: 600-1-198 CIP 1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 201-487-5800
;            TELEFAX: 201-343-1684
;  INFORMATION FOR SEQ ID NO: 15:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 50 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       HYPOTHETICAL: NO
;       FRAGMENT TYPE: internal
;       SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-052-089A-15

  Query Match              6.7%;  Score 152.5;  DB 4;  Length 50;
  Best Local Similarity   55.6%;  Pred. No. 1e-09;
  Matches   25;  Conservative    8;  Mismatches   11;  Indels    1;  Gaps    1;

Qy     275 DSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLK-HKTCPMCK 318
           | ||:|:: |:  | :||| |:| :|  ||||| |  ||||:||
Db       2 DVCAICLDEYEDGDKLRILPCSHAYHCKCVDPWLTKTKKTCPVCK 46
```

```
RESULT 8
US-09-046-894-44
; Sequence 44, Application US/09046894
; Patent No. 6190857
;  GENERAL INFORMATION:
;    APPLICANT:  Ralph, David
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, Mark S.
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  DIAGNOSIS OF DISEASE STATE USING mRNA
;    TITLE OF INVENTION:  PROFILES IN PERIPHERAL LEUKOCYTES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/046,894
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/041,576
;      FILING DATE:  24-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:014
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-09-046-894-44

  Query Match             6.4%;  Score 144.5;  DB 4;  Length 40;
  Best Local Similarity   56.1%;  Pred. No. 5.6e-09;
  Matches   23;  Conservative    8;  Mismatches    9;  Indels    1;  Gaps    1;

Qy     277 CAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMC 317
           ||:|:: |:  | :||| |:| :|  |||||| | ||||:|
Db       1 CAICLDEYEDGDKLRILPCSHAYHCKCVDPWLTK-KTCPVC 40


RESULT 9
US-09-325-932A-49
; Sequence 49, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant develo
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 49
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Pinus radiata
US-09-325-932A-49

  Query Match             5.8%;  Score 131.5;  DB 4;  Length 104;
  Best Local Similarity   29.4%;  Pred. No. 8.5e-07;
  Matches   30;  Conservative   18;  Mismatches   29;  Indels   25;  Gaps    4;

Qy     265 QGDKEIGPDGDSCAVCIELYKPNDLVRIL-TCNHIFHKTCVDPWLLKHKTCPMCKCDILK 323
           :| |:    | ||||:  ::  :::|:|  | | ||  |:| || || :||:|:
Db      17 KGSKQ----GLECAVCLSKFEDIEILRLLPKCRHAFHIDCIDYWLEKHSSCPLCR----- 67

Qy     324 ALGIEVDVEDGSVSLQVPVSNEISN-----SASSHEEDNRSE 360
                 || : | :  :|    | |  :|:  |
Db      68 ----------HKVSAEDPANFTYTNSMRLMSQQDMRQDSNLE 99


RESULT 10
US-08-841-349-9
; Sequence 9, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
;  APPLICANT: MISHRA, LOPA
;  TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
;  FILE REFERENCE: XX/PO4470US0
;  CURRENT APPLICATION NUMBER: US/08/841,349B
;  CURRENT FILING DATE: 1997-04-30
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 395
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-08-841-349-9

  Query Match             5.7%;  Score 130;  DB 2;  Length 395;
  Best Local Similarity   28.7%;  Pred. No. 1.3e-05;
  Matches   33;  Conservative   17;  Mismatches   49;  Indels   16;  Gaps    3;

Qy     220 AATVGY----FIFYSARRLRNARAQSRKQRQLKADA-----------KKAIGRLQLRTLK 264
           |  : |    |: | |    | |:|        |:: |           |::|  |    :
Db     276 AEAISYVDPQFLTYMALEERLAQAMETALAHLESLAVDVEVANPPASKESIDALPEILVT 335

Qy     265 QGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKC 319
           :   :| :   | :|   |   ::   | |:| ||| ||  || |  |||:|:|
Db     336 EDHGAVGQE-MCCPICCSEYVKGEVATELPCHHYFHKPCVSIWLQKSGTCPVCRC 389


RESULT 11
US-09-268-140-2
; Sequence 2, Application US/09268140
; Patent No. 6268176
; GENERAL INFORMATION:
;  APPLICANT:  Gemmill, Robert M.
;  APPLICANT:  Drabkin, Harry A.
;  TITLE OF INVENTION:  TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED
;  FILE REFERENCE:  93445-00004
;  CURRENT APPLICATION NUMBER: US/09/268,140
;  CURRENT FILING DATE:  2000-03-12
;  PRIOR APPLICATION NUMBER:  US 60/077,723
;  PRIOR FILING DATE:  1998-03-12
;  NUMBER OF SEQ ID NOS:  46
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 664
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-268-140-2

  Query Match             5.6%;  Score 127.5;  DB 4;  Length 664;
  Best Local Similarity   25.2%;  Pred. No. 6.2e-05;
  Matches   38;  Conservative   18;  Mismatches   44;  Indels   51;  Gaps    6;
```

```
Qy    275 DSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDG 334
          | ||:|   :  :   ||  ||| ||  |:  ||    |||||     : : || |::|
Db    545 DVCAICYHEFTTS--ARITPCNHYFHALCLRKWLYIQDTCPMCH----QKVYIEDDIKDN 598

Qy    335 SVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVNHE 394
          |          :||               ::|:        || |   ::  |:    : |
Db    599 S---------NVSN--------------NNGFI-------PPNETPEEAVREAAAESDRE 628

Qy    395 ANSVAVDVIPHVDNPTFEEDETPNQETAVRE 425
          |               |:| |   :   ||
Db    629 LN--------------EDDSTDCDDDVQRE 644


RESULT 12
US-09-325-932A-46
; Sequence 46, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant develo
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 46
;   LENGTH: 199
;   TYPE: PRT
;   ORGANISM: Pinus radiata
US-09-325-932A-46

  Query Match             5.0%;  Score 114.5;  DB 4;  Length 199;
  Best Local Similarity   44.2%;  Pred. No. 0.00021;
  Matches   19;  Conservative    7;  Mismatches   16;  Indels    1;  Gaps    1;

Qy    273 DGDSCAVCIELYKPNDLVRIL-TCNHIFHKTCVDPWLLKHKTC 314
          ||  ||||:  :: |:: |:|  | | ||  |:| |   | ||
Db    157 DGLECAVCLCEFEENEIARLLPNCRHNFHVECIDMWFRSHSTC 199


RESULT 13
US-08-699-103B-25
; Sequence 25, Application US/08699103B
; Patent No. 6107462
;  GENERAL INFORMATION:
;    APPLICANT:  Rine, Jasper D.
;    APPLICANT:  Hampton, Randolph
;    TITLE OF INVENTION:  GENES AND PROTEINS CONTROLLING
;    TITLE OF INVENTION:  CHOLESTEROL SYNTHESIS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  2200 Sand Hill Road, Suite 100
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/699,103B
;      FILING DATE:  16-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/002,581
;      FILING DATE:  17-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Grant D.
;      REGISTRATION NUMBER:  31,259
;      REFERENCE/DOCKET NUMBER:  09272/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/322-5070
;      TELEFAX:  650/854-0875
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  551 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  polypeptide
US-08-699-103B-25

  Query Match             4.8%;  Score 108.5;  DB 3;  Length 551;
  Best Local Similarity   20.4%;  Pred. No. 0.006;
  Matches   34;  Conservative   26;  Mismatches   58;  Indels   49;  Gaps    6;

Qy    273 DGDSCAVCIE--LYKPNDLV--------RILTCNHIFHKTCVDPWLLKHKTCPMCKCDIL 322
          | : | :|::  :: ||           : | | || | :|:  |: : :|||:|:  :
Db    345 DDNICIICMDELIHSPNQQTWKNKNKKPKRLPCGHILHLSCLKNWMERSQTCPICRLPVF 404

Qy    323 KALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQ 382
              | |:|     ||    : ::  :      |   |:|
Db    405 D--------EKGNVVQTTFTSNSDITTQTTVTDSTGIATDQQGFA--------------- 441

Qy    383 STNESLQLVNHEANSVAVDVIP-HVDNPTFEEDETPNQETAVREIKS 428
            ||            ||::|   :|      | | :|     :|
Db    442 --NE------------VDLLPTRTTSPDIRIVPTQNIDTLAMRTRS 473


RESULT 14
US-09-229-059-25
; Sequence 25, Application US/09229059
; Patent No. 6333172
;  GENERAL INFORMATION:
;    APPLICANT:  Rine, Jasper D.
;    APPLICANT:  Hampton, Randolph
;    TITLE OF INVENTION:  GENES AND PROTEINS CONTROLLING
;    TITLE OF INVENTION:  CHOLESTEROL SYNTHESIS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  2200 Sand Hill Road, Suite 100
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/229,059
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/699,103
;      FILING DATE:  16-AUG-1996
;      APPLICATION NUMBER:  60/002,581
;      FILING DATE:  17-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Grant D.
;      REGISTRATION NUMBER:  31,259
;      REFERENCE/DOCKET NUMBER:  09272/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/322-5070
;      TELEFAX:  650/854-0875
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  551 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  polypeptide
US-09-229-059-25
```

```
  Query Match             4.8%;  Score 108.5;  DB 4;  Length 551;
  Best Local Similarity   20.4%;  Pred. No. 0.006;
  Matches   34;  Conservative   26;  Mismatches   58;  Indels   49;  Gaps    6;

Qy     273 DGDSCAVCIE--LYKPNDLV--------RILTCNHIFHKTCVDPWLLKHKTCPMCKCDIL 322
           | : | :|::  :: ||           : | | || | :|:  |: : :|||:|:  :
Db     345 DDNICIICMDELIHSPNQQTWKNKNKKPKRLPCGHILHLSCLKNWMERSQTCPICRLPVF 404

Qy     323 KALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQ 382
                 | |:|      ||    : ::  :      |   |:|
Db     405 D--------EKGNVVQTTFTSNSDITTQTTVTDSTGIATDQQGFA--------------- 441

Qy     383 STNESLQLVNHEANSVAVDVIP-HVDNPTFEEDETPNQETAVREIKS 428
             ||             ||::|  :|       | | :|     :|
Db     442 --NE------------VDLLPTRTTSPDIRIVPTQNIDTLAMRTRS 473


RESULT 15
US-08-968-751-2
; Sequence 2, Application US/08968751
; Patent No. 5948643
;  GENERAL INFORMATION:
;    APPLICANT:  Rubinfeld, Bonnee
;    APPLICANT:  Polakis, Paul G.
;    APPLICANT:  Ligenfelter, Carol
;    APPLICANT:  Vuong, Terilyn T.
;    TITLE OF INVENTION:  MODULATORS OF BRCA1 ACTIVITY
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ONYX Pharmaceuticals, Inc.
;      STREET:  3031 Research Drive
;      CITY:  Richmond
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94806
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,751
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Giotta, Gregory
;      REGISTRATION NUMBER:  32,028
;      REFERENCE/DOCKET NUMBER:  ONYX1024 GG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 262-8710
;      TELEFAX:  (510) 222-9758
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 469 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-968-751-2

  Query Match             4.7%;  Score 107.5;  DB 2;  Length 469;
  Best Local Similarity   22.6%;  Pred. No. 0.0059;
  Matches   36;  Conservative   30;  Mismatches   82;  Indels   11;  Gaps    4;

Qy     277 CAVCIELYKPNDLVRILTCNHIFHKTCVDPWL--LKHKTCPMC-----KCDILKALGIEV 329
           | :| : :  :  |  : | | ||  |:  |       :||| |     |  |:  |  ::
Db       7 CTICSDFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSRTCPQCRIQVGKRTIINKLFFDL 66

Qy     330 DVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQ 389
             |: :|     : ||: |  :   : :: :  |     :  |    |||   :|  |||
Db      67 AQEEENVLDAEFLKNELDNVRAQLSQKDKEKRDSQ---VIIDTLRDTLEER-NATVVSLQ 122

Qy     390 LVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 428
```

```
               :|  :   :   :     ::|||   :   | ::|
Db     123 QALGKAEMLCSTLKKQMKYLEQQQDETKQAQEEARRLRS 161


Search completed: April 22, 2003, 08:27:49
Job time : 19 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 22, 2003, 08:16:45 ; Search time 76 Seconds
(without alignments)
750.411 Million cell updates/sec

Title: US-09-854-300-8
Perfect score: 2275
Sequence: 1 MGPPPGAGVSCRGGCGFSRL.....PTFEEDETPNQETAVREIKS 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2275	100.0	428	23	AAB47800	Human GRAIL. Homo
2	2255	99.1	428	21	AAY94897	Human protein clon
3	2255	99.1	428	23	AAB47874	ISIGP-4. Homo sap
4	2160	94.9	428	23	AAU74918	Mouse goliath prot
5	2115	93.0	428	23	AAU74927	Synthetic goliath
6	2111	92.8	428	23	AAU74922	Synthetic goliath
7	2111	92.8	428	23	AAU74925	Synthetic goliath
8	2110	92.7	428	23	AAU74923	Synthetic goliath
9	2108	92.7	428	23	AAU74926	Synthetic goliath
10	2106	92.6	428	23	AAU74924	Synthetic goliath
11	2090.5	91.9	427	23	AAB47799	Mouse GRAIL. Mus
12	1364	60.0	276	22	AAM40136	Human polypeptide
13	1134	49.8	225	22	AAU15897	Human novel secret
14	1027.5	45.2	294	22	AAU25459	Human mddt protein
15	838	36.8	165	22	AAU16353	Human novel secret
16	799	35.1	155	22	AAM41922	Human polypeptide
17	763.5	33.6	305	22	ABG05937	Novel human diagno
18	755.5	33.2	305	21	AAB25803	Human secreted pro
19	755.5	33.2	305	22	AAB75389	Human secreted pro
20	656.5	28.9	400	19	AAW63681	Human secreted pro
21	656.5	28.9	400	22	AAM93567	Human polypeptide,
22	650.5	28.6	400	22	AAM93452	Human polypeptide,
23	650.5	28.6	400	22	AAB88594	Human hydrophobic
24	649.5	28.5	400	23	ABB90388	Human polypeptide
25	617.5	27.1	419	22	AAB88558	Human hydrophobic
26	617.5	27.1	419	23	ABB77003	Human protein sequ
27	617.5	27.1	419	23	ABB77025	Human protein sequ
28	617.5	27.1	419	23	ABG64937	Human albumin fusi
29	617.5	27.1	419	23	ABG64938	Human albumin fusi
30	617.5	27.1	419	23	ABB90389	Human polypeptide
31	592.5	26.0	419	23	AAU74921	Mouse g1 protein s
32	586	25.8	462	23	AAU74919	Human goliath prot
33	551	24.2	115	23	ABB90387	Human polypeptide
34	545	24.0	341	22	ABB51081	Human secreted pro
35	537	23.6	312	22	ABB11857	Human secreted pro
36	537	23.6	312	22	AAM79358	Human protein SEQ
37	537	23.6	341	20	AAW89003	Polypeptide fragme
38	532.5	23.4	210	21	AAB41793	Human ORFX ORF1557
39	525	23.1	257	22	ABG12676	Novel human diagno
40	468.5	20.6	276	19	AAW69602	Human zinc binding
41	468.5	20.6	276	19	AAW63696	Human secreted pro
42	468.5	20.6	276	23	AAU74920	Human goliath-2 pr
43	462.5	20.3	276	20	AAW88996	Polypeptide fragme
44	462.5	20.3	276	22	ABB51074	Human secreted pro
45	438.5	19.3	167	20	AAW88995	Polypeptide fragme

ALIGNMENTS

RESULT 1
AAB47800
ID AAB47800 standard; Protein; 428 AA.
XX
AC AAB47800;
XX
DT 25-MAR-2002 (first entry)
XX
DE Human GRAIL.
XX
KW Murine; human; GRAIL; anergy; attenuation; tyrosine; phosphorylation;
KW antigenic stimulation; interleukin-2; gene therapy; polymorphism; IL-2;
KW autoimmune disease; tumour cell; cancer; transplant rejection.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 339..413
FT /label= Zinc_RING_finger_domain
FT /note= "Mediates protein/protein interaction"
XX
PN WO200185943-A1.
XX
PD 15-NOV-2001.
XX
PF 11-MAY-2001; 2001WO-US15385.
XX
PR 11-MAY-2000; 2000US-203513P.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Ford GS, Bloom D, Fathman CG;
XX

DR WPI; 2002-055597/07.
DR N-PSDB; AAI72095.
XX
PT Novel nucleic acid sequences of anergy associated genes, including
PT GRAIL gene useful in the evaluation of pathophysiology or immunotherapy
PT of cancer, autoimmune disease and transplant rejection -
XX
PS Claim 13; Page 47-48; 50pp; English.
XX
CC This sequence shows human GRAIL protein. The GRAIL gene is an
CC anergy associated gene which is upregulated during the early stages
CC of induction of anergy. GRAIL has been shown to attenuate IL-2
CC transcription in T-cells during response to antigenic stimulation.
CC GRAIL DNA is useful for decreasing the responsiveness of a T cell
CC population, especially synthesis of interleukin-2 (IL-2) in response to
CC antigenic stimulation, by up-regulating GRAIL activity in the T cell
CC population. GRAIL DNA is useful in producing compositions that modulate
CC induction or maintenance of anergy, for gene therapy, mapping
CC functional regions of the encoded protein, to analyse a patient sample
CC for the presence of polymorphisms or alterations in expression of
CC sequences associated with T cell anergy, disease states, genetic
CC predisposition to a disease state, and in studying associated
CC physiological pathways. Modulation of the gene activity in vivo is
CC useful for prophylactic and therapeutic purposes, such as treating
CC autoimmune disease and to enhance immune response to tumour cells, and
CC identification of anergic T cells. GRAIL DNA is useful in the
CC evaluation of the pathophysiology or immunotherapy of cancer,
CC autoimmune disease, and transplant rejection. Genetic sequences
CC involved in anergy induction are useful as markers in the evaluation of
CC specific immunotherapies. Functional characterization of genes involved
CC in anergy induction allows the elucidation of the mechanisms of T cell
CC anergy, including the transcriptional blockade of IL-2, which may be
CC manipulated to regulate T cell responses in human disease. Modulation
CC of expression of GRAIL gene is useful in manipulating the anergic
CC state. The genetic sequences find use alone or in combinations in
CC determining the expression profile of cells relating to anergy, for
CC example in screening of candidate biologically active compounds for
CC modulation of T cell anergy. Expression of GRAIL has been found to
CC block tyrosine phosphorylation of a protein present in T cells during
CC activation.
XX
SQ Sequence 428 AA;

Query Match 100.0%; Score 2275; DB 23; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.3e-236;
Matches 428; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420

Qy     421 TAVREIKS 428
          ||||||||
Db     421 TAVREIKS 428
```

RESULT 2
AAY94897
ID AAY94897 standard; Protein; 428 AA.
XX
AC AAY94897;
XX
DT 12-JUN-2000 (first entry)
XX
DE Human protein clone HP10574.
XX
KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW coagulation disorder; myocardial infarction; inflammatory condition;
KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW nephritis; therapy.
XX
OS Homo sapiens.
XX
PN WO200005367-A2.
XX
PD 03-FEB-2000.
XX
PF 22-JUL-1999; 99WO-JP03929.
XX
PR 24-JUL-1998; 98JP-0208820.
PR 07-AUG-1998; 98JP-0224105.
PR 25-AUG-1998; 98JP-0238116.
PR 09-SEP-1998; 98JP-0254736.
PR 29-SEP-1998; 98JP-0275505.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2000-182694/16.
XX
PT Novel human proteins having hydrophobic domains useful for treating
PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS Claim 1; Page 323-324; 351pp; English.
XX
CC This sequence represents a human protein of the invention, which has
CC hydrophobic domains. The DNA sequences can be used as a probe or as a
CC genetic marker. The protein can also be used as a marker, and to identify
CC potential genetic disorders. The DNA and protein can also be used as
CC nutritional sources or supplements. The protein exhibits cytokine, cell
CC proliferation, cell differentiation activities and induces production of
CC other cytokines in certain cell populations. The protein also exhibits
CC immune stimulating or immune suppressing activity. It can be used in the
CC treatment of various immune deficiencies and disorders, and to treat
CC infectious diseases caused by viral, bacterial, fungal or other
CC infections. The protein is also used for treating autoimmune disorders
CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC arthritis. It is also useful in the treatment of allergic reactions and
CC conditions such as asthma, and in immune suppression after organ
CC transplantation. The protein is useful in regulation of haematopoiesis
CC and consequently in the treatment of myeloid or lymphoid cell
CC deficiencies. It is also used in compositions for tissue growth or

```
CC   regeneration. The protein is also used in the treatment of osteoporosis
CC   or osteoarthritis and in the treatment of periodontal disease and other
CC   tooth repair processes. The protein is used in the treatment of nervous
CC   system disorders such as Alzheimer's disease, Parkinson's disease, and
CC   Huntington's disease. They are useful for protection or regeneration and
CC   treatment of lung or liver fibrosis, reperfusion injury in various
CC   tissues, and conditions resulting from systemic cytokine damage. They are
CC   also used for promoting or inhibiting tissue differentiation. They are
CC   also used as contraceptives since they exhibit activin or inhibin related
CC   activities and as a fertility inducing therapeutic. They are used for
CC   treating various coagulation disorders and in treatment and prevention of
CC   conditions resulting from coagulation activities e.g. myocardial
CC   infarction or stroke. They also acts as receptors, receptor ligands or
CC   inhibitors or agonists of receptor/ligand interactions. They are used to
CC   treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC   reperfusion injury, arthritis, and nephritis. They can be used to
CC   prevent tumours.
XX
SQ   Sequence   428 AA;

  Query Match             99.1%;  Score 2255;  DB 21;  Length 428;
  Best Local Similarity   99.3%;  Pred. No. 1.9e-234;
  Matches  425;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     121 QRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 TAVREIKS 428


RESULT 3
AAB47874
ID   AAB47874 standard; Protein; 428 AA.
XX
AC   AAB47874;
XX
DT   15-APR-2002  (first entry)
XX
DE   ISIGP-4.
XX
KW   Human; intracellular signalling protein; ISIGP; gene; autoimmune;
KW   cell proliferation; inflammation; gastrointestinal disorder;
KW   reproductive disorder; developmental disorder.
XX
OS   Homo sapiens.
XX

FH   Key             Location/Qualifiers
FT   Peptide         1..41
FT                   /label= Signal_peptide
FT                   /note= "identified using HMMER and SPSCAN"
FT   Modified-site   10
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   48
FT                   /note= "potentially N-glycosylated"
FT   Modified-site   50
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   59
FT                   /note= "potentially N-glycosylated"
FT   Modified-site   61
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   75
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   101
FT                   /note= "potentially N-glycosylated"
FT   Modified-site   127
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   151
FT                   /note= "potentially O-phosphorylated"
FT   Domain          206..229
FT                   /label= Transmembrane_domain
FT                   /note= "identified using HMMER"
FT   Modified-site   230
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   241
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   262
FT                   /note= "potentially O-phosphorylated"
FT   Domain          273..323
FT                   /label= Zinc_finger_C3CH4_type
FT                   /note= "identified using BLAST_DOMO"
FT   Domain          273..328
FT                   /label= Zinc_finger_C3CH4_type
FT                   /note= "identified using PROFILESCAN"
FT   Domain          277..317
FT                   /label= Zinc_finger_C3CH4_type
FT                   /note= "identified using HMMER_PFAM"
FT   Domain          292..306
FT                   /label= PHD_finger
FT                   /note= "identified using BLIMPS_PFAM"
FT   Modified-site   302
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   351
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   352
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   357
FT                   /note= "potentially N-glycosylated"
FT   Modified-site   383
FT                   /note= "potentially O-phosphorylated"
FT   Modified-site   385
FT                   /note= "potentially N-glycosylated"
FT   Modified-site   410
FT                   /note= "potentially O-phosphorylated"
XX
PN   WO200194391-A2.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-US18595.
XX
PR   08-JUN-2000; 2000US-210582P.
PR   16-JUN-2000; 2000US-212443P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yue H,  He A,  Nguyen DB,  Yao MG,  Bandman O,  Burford N,  Tang YT;
PI   Xu Y,  Hafalia A,  Azimzai Y,  Walia NK;
XX
DR   WPI; 2002-154564/20.
```

```
DR   N-PSDB; AAI72321.
XX
PT   New human intracellular signaling protein and polynucleotides useful
PT   for diagnosing, treating or preventing cell proliferative,
PT   autoimmune/inflammatory, gastrointestinal, reproductive and
PT   developmental disorders  -
XX
PS   Claim 1; Page 99-100; 106pp; English.
XX
CC   The sequences given in AAB47871-75 show novel human intracellular
CC   signalling proteins (ISIGP).  The polynucleotides and ISIGP proteins
CC   may be used for the diagnosis, treatment or prevention of cell
CC   proliferative, autoimmune/inflammatory, gastrointestinal, reproductive
CC   and developmental disorders. This protein has homology to the mouse
CC   gl-related zinc finger protein.
XX
SQ   Sequence   428 AA;

  Query Match             99.1%;  Score 2255;  DB 23;  Length 428;
  Best Local Similarity   99.3%;  Pred. No. 1.9e-234;
  Matches  425;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     121 QRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 TAVREIKS 428


RESULT 4
AAU74918
ID   AAU74918 standard; Protein; 428 AA.
XX
AC   AAU74918;
XX
DT   23-APR-2002  (first entry)
XX
DE   Mouse goliath protein sequence.
XX
KW   Mouse; goliath protein; antiangiogenic; vasotropic; gene therapy;
KW   dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW   cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW   unrecovered nerve trauma.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         20..32
FT                   /label= Signal_sequence
FT   Protein         33..428
FT                   /label= Mature_goliath_protein
FT   Domain          33..205
FT                   /label= Extracellular_N_terminal_domain
FT   Domain          206..225
FT                   /label= Transmembrane_domain
FT   Domain          226..428
FT                   /label= Intracellular_C_terminal_domain
FT   Domain          277..317
FT                   /label= C3H2C3_ring_finger_motif
XX
PN   WO200193681-A1.
XX
PD   13-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-US18000.
XX
PR   02-JUN-2000; 2000US-0586398.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Harland R,  Baker JC;
XX
DR   WPI; 2002-147637/19.
DR   N-PSDB; ABK12983.
XX
PT   New compositions comprising goliath proteins, useful for modulating
PT   angiogenesis or neurogenesis in mammals e.g. for preventing or
PT   treating undesirable vascularisation of a tumour, ischaemia or
PT   neurodegenerative disease  -
XX
PS   Claim 1; Page 29-30; 45pp; English.
XX
CC   The present invention relates to a new pharmaceutical composition that
CC   comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC   The present amino acid sequence represents the mouse goliath protein
CC   that was used in the invention for modulating angiogenesis or
CC   neurogenesis.
XX
SQ   Sequence   428 AA;

  Query Match             94.9%;  Score 2160;  DB 23;  Length 428;
  Best Local Similarity   94.6%;  Pred. No. 3.5e-224;
  Matches  405;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||||| ||||||||||||||||||||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||||| || |||||||||||||||||||| ||||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
```

```
Db     241 SRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:|||||||||||||||||||||||||||||||||| ||:|| ||||:|||
Db     301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTASPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||||  |||||||| || ||:|||||||||||||:|||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVDVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428
```

RESULT 5
AAU74927
ID AAU74927 standard; Protein; 428 AA.
XX
AC AAU74927;
XX
DT 23-APR-2002 (first entry)
XX
DE Synthetic goliath protein sequence #6.
XX
KW Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW unrecovered nerve trauma.
XX
OS Synthetic.
XX
PN WO200193681-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-US18000.
XX
PR 02-JUN-2000; 2000US-0586398.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Harland R, Baker JC;
XX
DR WPI; 2002-147637/19.
XX
PT New compositions comprising goliath proteins, useful for modulating
PT angiogenesis or neurogenesis in mammals e.g. for preventing or
PT treating undesirable vascularisation of a tumour, ischaemia or
PT neurodegenerative disease -
XX
PS Disclosure; Page 37-38; 45pp; English.
XX
CC The present invention relates to a new pharmaceutical composition that
CC comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC has a sequence identity of at least 75% to the protein sequences
CC (AAU74918-AAU74921) fully defined in the specification. The composition
CC is useful for modulating angiogenesis or neurogenesis in mammals,
CC particularly in humans or mice. Specifically, the composition is useful
CC for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC The present amino acid sequence represents synthetic goliath protein #6
CC that is one of several artificial goliath proteins (AAU74922-AAU74927)
CC used in the invention for modulating angiogenesis or neurogenesis.
XX
SQ Sequence 428 AA;

Query Match 93.0%; Score 2115; DB 23; Length 428;
Best Local Similarity 92.8%; Pred. No. 2.6e-219;
Matches 397; Conservative 11; Mismatches 20; Indels 0; Gaps 0;

```
Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||| |||||||||||||||| ||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVGWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHHNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||| || |||||||||| |||||||||| ||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGIRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGKWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           ||||||||| |||||:|||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADLKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           ||||||||:|:|||||||||||||||||||||||||||||||| ||:|: ||||:|||
Db     301 MTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTANPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||||  |||||||| || ||:|||||||||||||  |:|||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVQVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428
```

RESULT 6
AAU74922
ID AAU74922 standard; Protein; 428 AA.
XX
AC AAU74922;
XX
DT 23-APR-2002 (first entry)
XX
DE Synthetic goliath protein sequence #1.
XX
KW Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW unrecovered nerve trauma.
XX
OS Synthetic.
XX
PN WO200193681-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-US18000.
XX
PR 02-JUN-2000; 2000US-0586398.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Harland R, Baker JC;
XX
DR WPI; 2002-147637/19.
XX
PT New compositions comprising goliath proteins, useful for modulating
PT angiogenesis or neurogenesis in mammals e.g. for preventing or
PT treating undesirable vascularisation of a tumour, ischaemia or
PT neurodegenerative disease -
XX
PS Disclosure; Page 32-33; 45pp; English.
XX
CC The present invention relates to a new pharmaceutical composition that
CC comprises a goliath polypeptide in dosage form. The goliath polypeptide

```
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC   The present amino acid sequence represents synthetic goliath protein #1
CC   that is one of several artificial goliath proteins (AAU74922-AAU74927)
CC   used in the invention for modulating angiogenesis or neurogenesis.
XX
SQ   Sequence   428 AA;

  Query Match             92.8%;  Score 2111;  DB 23;  Length 428;
  Best Local Similarity   92.8%;  Pred. No. 6.9e-219;
  Matches  397;  Conservative   10;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||||| ||||||||||||||||| ||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVHWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHINFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||| || |||||||||| |||||||||| ||||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGKRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGLWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           |||||||||| |||||:|||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADMKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           ||||||||:|:|||||||||||||||||||||||||||||||||||| ||:|  ||||:|||
Db     301 NTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTAQPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||| |||||||| || ||:|||||||||||| |:||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVRVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428


RESULT 7
AAU74925
ID   AAU74925 standard; Protein; 428 AA.
XX
AC   AAU74925;
XX
DT   23-APR-2002  (first entry)
XX
DE   Synthetic goliath protein sequence #4.
XX
KW   Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW   dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW   cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW   unrecovered nerve trauma.
XX
OS   Synthetic.
XX
PN   WO200193681-A1.
XX
PD   13-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-US18000.
XX
PR   02-JUN-2000; 2000US-0586398.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Harland R,  Baker JC;
XX
DR   WPI; 2002-147637/19.
XX
PT   New compositions comprising goliath proteins, useful for modulating
PT   angiogenesis or neurogenesis in mammals e.g. for preventing or
PT   treating undesirable vascularisation of a tumour, ischaemia or
PT   neurodegenerative disease  -
XX
PS   Disclosure; Page 35-36; 45pp; English.
XX
CC   The present invention relates to a new pharmaceutical composition that
CC   comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC   The present amino acid sequence represents synthetic goliath protein #4
CC   that is one of several artificial goliath proteins (AAU74922-AAU74927)
CC   used in the invention for modulating angiogenesis or neurogenesis.
XX
SQ   Sequence   428 AA;

  Query Match             92.8%;  Score 2111;  DB 23;  Length 428;
  Best Local Similarity   92.8%;  Pred. No. 6.9e-219;
  Matches  397;  Conservative   11;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||||| ||||||||||||||||| ||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVEWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHFNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||| || |||||||||| |||||||||| ||||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGGRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGHWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           |||||||||| |||||:|||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADIKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           ||||||||:|:|||||||||||||||||||||||||||||||||||| ||:|  ||||:|||
Db     301 LTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTAMPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||| |||||||| || ||:|||||||||||| |:||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVNVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428


RESULT 8
AAU74923
ID   AAU74923 standard; Protein; 428 AA.
XX
```

```
AC   AAU74923;
XX
DT   23-APR-2002  (first entry)
XX
DE   Synthetic goliath protein sequence #2.
XX
KW   Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW   dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW   cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW   unrecovered nerve trauma.
XX
OS   Synthetic.
XX
PN   WO200193681-A1.
XX
PD   13-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-US18000.
XX
PR   02-JUN-2000; 2000US-0586398.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Harland R,  Baker JC;
XX
DR   WPI; 2002-147637/19.
XX
PT   New compositions comprising goliath proteins, useful for modulating
PT   angiogenesis or neurogenesis in mammals e.g. for preventing or
PT   treating undesirable vascularisation of a tumour, ischaemia or
PT   neurodegenerative disease  -
XX
PS   Disclosure; Page 33-34; 45pp; English.
XX
CC   The present invention relates to a new pharmaceutical composition that
CC   comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC   The present amino acid sequence represents synthetic goliath protein #2
CC   that is one of several artificial goliath proteins (AAU74922-AAU74927)
CC   used in the invention for modulating angiogenesis or neurogenesis.
XX
SQ   Sequence   428 AA;

  Query Match             92.7%;  Score 2110;  DB 23;  Length 428;
  Best Local Similarity   92.8%;  Pred. No. 8.9e-219;
  Matches  397;  Conservative   11;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||||| ||||||||||||||||:||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVAWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           |||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHDNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||| || |||||||||| |||||||||||| ||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGERNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGFWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           |||||||||| ||||:|||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADGKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           ||||||||:|:|||||||||||||||||||||||||||||||||| ||:|  ||||:|||
Db     301 GTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTAIPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||| |||||||| || ||:|||||||||||||| |:|||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVKVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428


RESULT 9
AAU74926
ID   AAU74926 standard; Protein; 428 AA.
XX
AC   AAU74926;
XX
DT   23-APR-2002  (first entry)
XX
DE   Synthetic goliath protein sequence #5.
XX
KW   Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW   dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW   cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW   unrecovered nerve trauma.
XX
OS   Synthetic.
XX
PN   WO200193681-A1.
XX
PD   13-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-US18000.
XX
PR   02-JUN-2000; 2000US-0586398.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Harland R,  Baker JC;
XX
DR   WPI; 2002-147637/19.
XX
PT   New compositions comprising goliath proteins, useful for modulating
PT   angiogenesis or neurogenesis in mammals e.g. for preventing or
PT   treating undesirable vascularisation of a tumour, ischaemia or
PT   neurodegenerative disease  -
XX
PS   Disclosure; Page 36-37; 45pp; English.
XX
CC   The present invention relates to a new pharmaceutical composition that
CC   comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC   has a sequence identity of at least 75% to the protein sequences
CC   (AAU74918-AAU74921) fully defined in the specification. The composition
CC   is useful for modulating angiogenesis or neurogenesis in mammals,
CC   particularly in humans or mice. Specifically, the composition is useful
CC   for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC   e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC   e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC   neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC   The present amino acid sequence represents synthetic goliath protein #5
CC   that is one of several artificial goliath proteins (AAU74922-AAU74927)
CC   used in the invention for modulating angiogenesis or neurogenesis.
XX
SQ   Sequence   428 AA;

  Query Match             92.7%;  Score 2108;  DB 23;  Length 428;
  Best Local Similarity   92.8%;  Pred. No. 1.5e-218;
  Matches  397;  Conservative   11;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
```

```
          |||||| || |||||| :|||||||||||||| +|||||||||||||||| ||||||||||
Db      1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVFWRVPHTGVNR 60

Qy     61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
          ||||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||
Db     61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHGNFTVPTVWGSTVQVSWLALI 120

Qy    121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
          |||||||||||||| || |||||||||| |||||||||| ||||||||||||||||
Db    121 QRGGGCTFADKIHLASERGASGAVIFNFPGHRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy    181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    181 QSIQRGIQVTMVIEVGKKHGIWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy    241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
          |||||||||| |||||:|||||||||||||||||||||||||||||||||||||||||||
Db    241 SRKQRQLKADKKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy    301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
          ||||||||:|:|||||||||||||||||||||||||||||||||| ||:|  ||||:|||
Db    301 LTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTAMPHEEDSRSE 360

Qy    361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
          |||||||||| |||||||| || ||:|||||||||||||:|:|||||||||||||||:||
Db    361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVNVVPHVDNPTFEEDETPDQE 420

Qy    421 TAVREIKS 428
          ||||||| 
Db    421 AAVREIKS 428
```

RESULT 10
AAU74924
ID AAU74924 standard; Protein; 428 AA.
XX
AC AAU74924;
XX
DT 23-APR-2002 (first entry)
XX
DE Synthetic goliath protein sequence #3.
XX
KW Goliath protein; antiangiogenic; vasotropic; gene therapy;
KW dosage form; angiogenesis; neurogenesis; tumour; vascularisation;
KW cancer; ischaemia; neuroblastoma; neurodegenerative disease;
KW unrecovered nerve trauma.
XX
OS Synthetic.
XX
PN WO200193681-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-US18000.
XX
PR 02-JUN-2000; 2000US-0586398.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Harland R, Baker JC;
XX
DR WPI; 2002-147637/19.
XX
PT New compositions comprising goliath proteins, useful for modulating
PT angiogenesis or neurogenesis in mammals e.g. for preventing or
PT treating undesirable vascularisation of a tumour, ischaemia or
PT neurodegenerative disease -
XX
PS Disclosure; Page 34-35; 45pp; English.
XX
CC The present invention relates to a new pharmaceutical composition that
CC comprises a goliath polypeptide in dosage form. The goliath polypeptide
CC has a sequence identity of at least 75% to the protein sequences
CC (AAU74918-AAU74921) fully defined in the specification. The composition
CC is useful for modulating angiogenesis or neurogenesis in mammals,
CC particularly in humans or mice. Specifically, the composition is useful
CC for the prophylactic and/or therapeutic treatment of excess angiogenesis
CC e.g. undesirable vascularisation of a tumour or insufficient angiogenesis
CC e.g. ischaemia, excess neurogenesis e.g. neuroblastoma or insufficient
CC neurogenesis e.g. neurodegenerative disease or unrecovered nerve trauma.
CC The present amino acid sequence represents synthetic goliath protein #3
CC that is one of several artificial goliath proteins (AAU74922-AAU74927)
CC used in the invention for modulating angiogenesis or neurogenesis.
XX
SQ Sequence 428 AA;

```
  Query Match             92.6%;  Score 2106;  DB 23;  Length 428;
  Best Local Similarity   92.8%;  Pred. No. 2.4e-218;
  Matches  397;  Conservative   10;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
          |||||| || |||||| :|||||||||||| |||||||||||||||| ||||||||||
Db      1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVDWRVPHTGVNR 60

Qy     61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
          ||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||
Db     61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHENFTVPTVWGSTVQVSWLALI 120

Qy    121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
          |||||||||||||| || |||||||||| |||||||||| ||||||||||||||||
Db    121 QRGGGCTFADKIHLASERGASGAVIFNFPGFRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy    181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    181 QSIQRGIQVTMVIEVGKKHGGWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy    241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
          |||||||||| |||||:|||||||||||||||||||||||||||||||||||||||||||
Db    241 SRKQRQLKADHKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy    301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
          ||||||||:|:|||||||||||||||||||||||||||||||||| ||:|  ||||:|||
Db    301 ITCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTAKPHEEDSRSE 360

Qy    361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
          |||||||||| |||||||| || ||:|||||||||||||:|:|||||||||||||||:||
Db    361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVLVVPHVDNPTFEEDETPDQE 420

Qy    421 TAVREIKS 428
          ||||||| 
Db    421 AAVREIKS 428
```

RESULT 11
AAB47799
ID AAB47799 standard; Protein; 427 AA.
XX
AC AAB47799;
XX
DT 25-MAR-2002 (first entry)
XX
DE Mouse GRAIL.
XX
KW Murine; human; GRAIL; anergy; attenuation; tyrosine; phosphorylation;
KW antigenic stimulation; interleukin-2; gene therapy; polymorphism; IL-2;
KW autoimmune disease; tumour cell; cancer; transplant rejection.
XX
OS Mus musculus.
XX
FH Key   Location/Qualifiers
FT Domain   339..413
FT   /label= Zinc_RING_finger_domain
FT   /note= "Mediates protein/protein interaction"
XX
PN WO200185943-A1.
XX

PD 15-NOV-2001.
XX
PF 11-MAY-2001; 2001WO-US15385.
XX
PR 11-MAY-2000; 2000US-203513P.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Ford GS, Bloom D, Fathman CG;
XX
DR WPI; 2002-055597/07.
DR N-PSDB; AAI72094.
XX
PT Novel nucleic acid sequences of anergy associated genes, including
PT GRAIL gene useful in the evaluation of pathophysiology or immunotherapy
PT of cancer, autoimmune disease and transplant rejection -
XX
PS Claim 13; Page 44-45; 50pp; English.
XX
CC This sequence shows murine GRAIL protein. The GRAIL gene is an
CC anergy associated gene which is upregulated during the early stages
CC of induction of anergy. GRAIL has been shown to attenuate IL-2
CC transcription in T-cells during response to antigenic stimulation.
CC GRAIL DNA is useful for decreasing the responsiveness of a T cell
CC population, especially synthesis of interleukin-2 (IL-2) in response to
CC antigenic stimulation, by up-regulating GRAIL activity in the T cell
CC population. GRAIL DNA is useful in producing compositions that modulate
CC induction or maintenance of anergy, for gene therapy, mapping
CC functional regions of the encoded protein, to analyse a patient sample
CC for the presence of polymorphisms or alterations in expression of
CC sequences associated with T cell anergy, disease states, genetic
CC predisposition to a disease state, and in studying associated
CC physiological pathways. Modulation of the gene activity in vivo is
CC useful for prophylactic and therapeutic purposes, such as treating
CC autoimmune disease and to enhance immune response to tumour cells, and
CC identification of anergic T cells. GRAIL DNA is useful in the
CC evaluation of the pathophysiology or immunotherapy of cancer,
CC autoimmune disease, and transplant rejection. Genetic sequences
CC involved in anergy induction are useful as markers in the evaluation of
CC specific immunotherapies. Functional characterization of genes involved
CC in anergy induction allows the elucidation of the mechanisms of T cell
CC anergy, including the transcriptional blockade of IL-2, which may be
CC manipulated to regulate T cell responses in human disease. Modulation
CC of expression of GRAIL gene is useful in manipulating the anergic
CC state. The genetic sequences find use alone or in combinations in
CC determining the expression profile of cells relating to anergy, for
CC example in screening of candidate biologically active compounds for
CC modulation of T cell anergy. Expression of GRAIL has been found to
CC block tyrosine phosphorylation of a protein present in T cells during
CC activation.
XX
SQ Sequence 427 AA;

Query Match 91.9%; Score 2090.5; DB 23; Length 427;
Best Local Similarity 92.3%; Pred. No. 1.1e-216;
Matches 395; Conservative 11; Mismatches 21; Indels 1; Gaps 1;

```
Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||||| |||||||||||||||||||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVSWRVPHTGVTA 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
                |||||||||:||:|||||||||||||||||||||||||||||||||||||||||||
Db      61 RC-GAEREGVYGQDSPLKPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 119

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           ||||||||||||||| || |||||||||||||||||||||||| |||||||||||||||||
Db     120 QRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 179

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 239

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db     240 SRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 299

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:|||||||||||||||||||||||||||||||||| ||:|| ||||:|||
Db     300 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTASPHEEDSRSE 359

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||||| |||||||| || ||:|||||||||||||||:|||||||||||||||:||
Db     360 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVDVVPHVDNPTFEEDETPDQE 419

Qy     421 TAVREIKS 428
           ||||||||
Db     420 AAVREIKS 427
```

RESULT 12
AAM40136
ID AAM40136 standard; Protein; 276 AA.
XX
AC AAM40136;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3281.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59292.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 5; SEQ ID NO 3281; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

```
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   276 AA;

  Query Match             60.0%;  Score 1364;  DB 22;  Length 276;
  Best Local Similarity   99.2%;  Pred. No. 2e-138;
  Matches  263;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     164 VDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATV 223
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12 VDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATV 71

Qy     224 GYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIEL 283
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 GYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIEL 131

Qy     284 YKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVS 343
           |||||||||||||||||||||||:|:||||||||||||||||||||||||||||||||||||
Db     132 YKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVS 191

Qy     344 NEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVI 403
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     192 NEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVI 251

Qy     404 PHVDNPTFEEDETPNQETAVREIKS 428
           |||||||||||||||||||||||||
Db     252 PHVDNPTFEEDETPNQETAVREIKS 276


RESULT 13
AAU15897
ID   AAU15897 standard; Protein; 225 AA.
XX
AC   AAU15897;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 850.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
```

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS25884.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 850; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 49.8%; Score 1134; DB 22; Length 225;
Best Local Similarity 99.1%; Pred. No. 1.1e-113;
Matches 218; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy    209 FFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDK 268
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 FFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDK 65

Qy    269 EIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIE 328
          ||||||||||||||||||||||||||||||||||||||:|:||||||||||||||||||||
Db     66 EIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIE 125

Qy    329 VDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESL 388
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 VDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESL 185

Qy    389 QLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 428
          |||||||||||||||||||||||||||||||||||||||||
Db    186 QLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 225
```

RESULT 14
AAU25459
ID AAU25459 standard; Protein; 294 AA.
XX
AC AAU25459;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human mddt protein from clone LI:757439.1:2000MAY01.
XX
KW Human; molecules for disease detection and treatment; mddt;
KW Antiarteriosclerotic; hepatotropic; antipsoriatic; cytostatic;
KW immunosuppressive; antidiabetic; antiasthmatic; neuroprotective;
KW osteopathic; antiarthritic; cell proliferative disorder;
KW arteriosclerosis; cirrhosis; psoriasis; cancer; adenocarcinoma;
KW leukaemia; breast cancer; autoimmune disorder; AIDS;
KW acquired immunodeficiency syndrome; Addison's disease;
KW diabetes mellitus; asthma; multiple sclerosis; osteoarthritis.
XX
OS Homo sapiens.
XX
PN WO200162922-A2.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US05896.
XX
PR 24-FEB-2000; 2000US-0185213.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.

```
PR   17-MAY-2000; 2000US-0205323.
PR   17-MAY-2000; 2000US-0205324.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI   Chen A, D'Sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;
PI   Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
PI   Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI   Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI   Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR   WPI; 2001-570631/64.
DR   N-PSDB; AAS42511.
XX
PT   New disease detection and treatment molecule polynucleotides and
PT   polypeptides, useful for diagnosis and treatment of arteriosclerosis,
PT   cirrhosis, psoriasis, cancer, autoimmune disorders, diabetes mellitus,
PT   asthma and multiple sclerosis -
XX
PS   Claim 27; Page 170-171; 183pp; English.
XX
CC   The invention relates to novel human molecules for disease
CC   detection and treatment (mddt proteins) and the polynucleotides encoding
CC   them. The MDDT polynucleotides and polypeptides are useful for diagnostic
CC   and therapeutic purposes e.g. to diagnose and treat cell proliferative
CC   disorders (e.g. arteriosclerosis, cirrhosis and psoriasis) cancers (e.g.
CC   adenocarcinoma, leukaemia and breast cancer) autoimmune disorders
CC   (e.g. acquired immunodeficiency syndrome (AIDS) and Addison's disease)
CC   diabetes mellitus, asthma, multiple sclerosis, osteoarthritis, and many
CC   more diseases given in the specification. The present sequence
CC   represents an mddt protein of the invention.
XX
SQ   Sequence   294 AA;

  Query Match             45.2%;  Score 1027.5;  DB 22;  Length 294;
  Best Local Similarity   76.1%;  Pred. No. 5.5e-102;
  Matches  201;  Conservative    7;  Mismatches   25;  Indels   31;  Gaps    3;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 102

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||| ||
Db     103 TVWELSEEGVYGPDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLGLI 162

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNL---KGT 177
           |||||||||||||||| |||||||||||||||||||||||||||||      | :
Db     163 QRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIRQSERHKNS 222

Qy     178 KILQSIQRGIQVTMVIEVGKKHGPWVNH------------YSIFFVSVSFFIITAATVG 224
           |  |          |:|     |                :: |    | ::  ||||
Db     223 AI--------------YSKRHTSDNGHRSREKTWPLGESLFNFFRFLCPFLLLRRATVG 267

Qy     225 YFIFYSARRLRNARAQSRKQRQLK 248
           |||||||||||||||||||| :|
Db     268 YFIFYSARRLRNARAQSRKQRPIK 291


RESULT 15
AAU16353
ID   AAU16353 standard; Protein; 165 AA.
XX
AC   AAU16353;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1306.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
```

PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26340.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 1306; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 36.8%; Score 838; DB 22; Length 165;
Best Local Similarity 97.5%; Pred. No. 6.6e-82;
Matches 157; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy    268 KEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGI 327
          : |||||||||||||||||||||||||||||||||||||||||||:|:|||||||||||||||
Db      5 RXIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGI 64

Qy    328 EVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNES 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 EVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNES 124

Qy    388 LQLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 428
          |||||||||||||||||||||||||||||||||||||||||||
Db    125 LQLVNHEANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 165
```

Search completed: April 22, 2003, 08:25:54
Job time : 78 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 22, 2003, 08:22:11 ; Search time 35 Seconds
(without alignments)
2519.662 Million cell updates/sec

Title: US-09-854-300-8
Perfect score: 2275
Sequence: 1 MGPPPGAGVSCRGGCGFSRL.....PTFEEDETPNQETAVREIKS 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2255	99.1	428	4	Q8TEB7	Q8teb7 homo sapien
2	2213	97.3	428	4	Q96RF3	Q96rf3 homo sapien
3	2160	94.9	428	11	Q9D304	Q9d304 mus musculu
4	2159	94.9	428	11	Q9JJF8	Q9jjf8 mus musculu
5	2149	94.5	428	11	Q9DBN3	Q9dbn3 mus musculu
6	1218	53.5	237	11	Q9CVG1	Q9cvg1 mus musculu
7	930	40.9	376	6	Q95JW9	Q95jw9 macaca fasc
8	927	40.7	376	6	Q95K04	Q95k04 macaca fasc
9	910.5	40.0	376	4	Q8WVZ7	Q8wvz7 homo sapien
10	829	36.4	316	11	Q9D3V4	Q9d3v4 mus musculu
11	714	31.4	156	4	Q9H5E4	Q9h5e4 homo sapien
12	650.5	28.6	426	4	Q8WUU3	Q8wuu3 homo sapien
13	595.5	26.2	419	11	Q8VEM1	Q8vem1 mus musculu
14	592.5	26.0	419	11	Q9QZQ6	Q9qzq6 mus musculu
15	586	25.8	462	4	Q9ULK6	Q9ulk6 homo sapien
16	468.5	20.6	276	4	Q9P0J9	Q9p0j9 homo sapien
17	412.5	18.1	461	5	Q8T0M7	Q8t0m7 drosophila
18	356.5	15.7	180	11	Q9CY01	Q9cy01 mus musculu
19	346	15.2	286	5	Q9W0W3	Q9w0w3 drosophila
20	327	14.4	473	5	O76671	O76671 caenorhabdi
21	227.5	10.0	310	10	Q9M622	Q9m622 arabidopsis
22	212	9.3	347	11	Q91XF4	Q91xf4 mus musculu
23	211.5	9.3	783	4	Q9NXD0	Q9nxd0 homo sapien
24	210	9.2	350	4	Q9H6Y7	Q9h6y7 homo sapien
25	210	9.2	536	5	Q9VI20	Q9vi20 drosophila
26	208	9.1	537	10	Q8S7T8	Q8s7t8 oryza sativ
27	205	9.0	424	10	Q9C8W4	Q9c8w4 arabidopsis
28	205	9.0	448	10	Q8VZ14	Q8vz14 arabidopsis
29	204.5	9.0	431	10	Q9M0N7	Q9m0n7 arabidopsis
30	203	8.9	444	10	Q9M621	Q9m621 arabidopsis
31	203	8.9	456	5	Q9XX98	Q9xx98 caenorhabdi
32	194.5	8.5	324	4	Q9Y3V1	Q9y3v1 homo sapien
33	191	8.4	891	4	Q9ULT6	Q9ult6 homo sapien
34	190	8.4	378	10	O64763	O64763 arabidopsis
35	189.5	8.3	354	10	Q8S7D2	Q8s7d2 oryza sativ
36	185	8.1	332	10	Q9LQM2	Q9lqm2 arabidopsis
37	184.5	8.1	508	10	Q9C804	Q9c804 arabidopsis
38	184	8.1	404	10	Q9C7T5	Q9c7t5 arabidopsis
39	182.5	8.0	226	11	Q9CYM6	Q9cym6 mus musculu
40	177	7.8	420	4	O75866	O75866 homo sapien
41	177	7.8	420	4	Q8WWF5	Q8wwf5 homo sapien
42	176.5	7.8	362	10	Q9LF64	Q9lf64 arabidopsis
43	172	7.6	291	10	Q9SRQ8	Q9srq8 arabidopsis
44	172	7.6	322	10	Q9SMQ0	Q9smq0 arabidopsis
45	172	7.6	356	10	Q9SS59	Q9ss59 arabidopsis

ALIGNMENTS

RESULT 1
Q8TEB7
ID Q8TEB7 PRELIMINARY; PRT; 428 AA.
AC Q8TEB7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ23684 fis, clone HEP09821.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK074264; BAB85033.1; -.
SQ SEQUENCE 428 AA; 46520 MW; 32F9CDB32BF208FA CRC64;

Query Match 99.1%; Score 2255; DB 4; Length 428;
Best Local Similarity 99.3%; Pred. No. 9.3e-196;
Matches 425; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     121 QRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
```

```
Db    181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy    241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy    301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
          |||||||||:|:||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360

Qy    361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420

Qy    421 TAVREIKS 428
          ||||||||
Db    421 TAVREIKS 428


RESULT 2
Q96RF3
ID   Q96RF3      PRELIMINARY;      PRT;   428 AA.
AC   Q96RF3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   GRAIL.
GN   GRAIL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Soares L.R., Fathman G.;
RT   "Grail: a novel ring finger protein upregulated in anergic T cells.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF394689; AAK77554.1; -.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
KW   Zinc-finger.
SQ   SEQUENCE   428 AA;  46491 MW;  E003DBED8968E45A CRC64;

  Query Match             97.3%;  Score 2213;  DB 4;  Length 428;
  Best Local Similarity   97.7%;  Pred. No. 5.9e-192;
  Matches  418;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGPLPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy     61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy    121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    121 QRGGGCTFADKIHLAYERGASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180

Qy    181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
          |||||||||||||||||||||||||||||| ||||| || |||||||| ||||||| |
Db    181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVPFFIITXATGGYFIFYSAXRLRNARXQ 240

Qy    241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 SXKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy    301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
          ||'|||||:|:||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360

Qy    361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420

Qy    421 TAVREIKS 428
          ||||||||
Db    421 TAVREIKS 428


RESULT 3
Q9D304
ID   Q9D304      PRELIMINARY;      PRT;   428 AA.
AC   Q9D304;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   1300002C13Rik protein (RIKEN cDNA 1300002C13 gene) (Zinc ring
DE   finger-containing protein GRAIL).
GN   1300002C13RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=CECUM;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST TUMOR;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2;
RA   Ford G.S., Bloom D., Paragas V., Anandasabapathy N., Skrenta H.,
RA   Ermann J., Cron R.Q., Lewis D.B., Fathman G.;
RT   "GRAIL: A novel gene expressed in anergic T cells that inhibits
RT   NFAT/AP-1 activation and IL-2 transcription.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK018582; BAB31291.1; -.
DR   EMBL; BC010477; AAH10477.1; -.
DR   EMBL; AF426411; AAL34514.1; -.
DR   MGD; MGI:1914139; 1300002C13Rik.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
KW   Zinc-finger.
```

```
SQ   SEQUENCE   428 AA;  46276 MW;  480DCF46C75E238F CRC64;

  Query Match             94.9%;  Score 2160;  DB 11;  Length 428;
  Best Local Similarity   94.6%;  Pred. No. 3.8e-187;
  Matches  405;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||| ||||||||||||||||||||||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||| || |||||||||||||||||||| ||||||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:|||||||||||||||||||||||||||||||| ||:|| ||||:|||
Db     301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTASPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||||| |||||||| || ||:|||||||||||||:|||||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVDVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428


RESULT 4
Q9JJF8
ID   Q9JJF8      PRELIMINARY;      PRT;   428 AA.
AC   Q9JJF8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Brain cDNA, clone MNCb-3816, similar to AF171875 g1-related zinc
DE   finger protein (Mus musculus).
GN   1300002C13RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL;
RA   Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA   Hashimoto K.;
RT   "Isolation of full-length cDNA clones from mouse brain cDNA library
RT   made by oligo-capping method.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AB041548; BAA95033.1; -.
DR   MGD; MGI:1914139; 1300002C13Rik.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
KW   Zinc-finger.
SQ   SEQUENCE   428 AA;  46262 MW;  E13A6DCAD67DECBD CRC64;

  Query Match             94.9%;  Score 2159;  DB 11;  Length 428;
  Best Local Similarity   94.4%;  Pred. No. 4.6e-187;
  Matches  404;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||| ||||||||||||||||||||||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||:||| || |||||||||||||||||||| ||||||||||||||||||||
Db     121 QRGGGCTFADKVHLASERGASGAVIFNFPGTRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:|||||||||||||||||||||||||||||||| ||:|| ||||:|||
Db     301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTASPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||||| |||||||| || ||:|||||||||||||:|||||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVDVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428


RESULT 5
Q9DBN3
ID   Q9DBN3      PRELIMINARY;      PRT;   428 AA.
AC   Q9DBN3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   1300002C13Rik protein.
GN   1300002C13RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
```

```
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK004847; BAB23613.1; -.
DR   MGD; MGI:1914139; 1300002C13Rik.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
KW   Zinc-finger.
SQ   SEQUENCE   428 AA;  46292 MW;  EBA2B505218288B8 CRC64;

  Query Match             94.5%;  Score 2149;  DB 11;  Length 428;
  Best Local Similarity   94.2%;  Pred. No. 3.7e-186;
  Matches  403;  Conservative   10;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||||| |||||||||||||||||||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||| ||||| |||| || |||||||||||||+|||||||||| ||||||||||||||||||
Db     121 QRGGSCTFADNIHLASERGASGAVIFNFPGTRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240

Qy     241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
           |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db     241 SRKQRQLKADAKKAIGKLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300

Qy     301 KTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           |||||||||:|:|||||||||||||||||||||||||||||||||| ||:|| ||||:|||
Db     301 KTCVDPWLLEHRTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEASNTASPHEEDSRSE 360

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           |||||||||||| |||||||| || ||:|||||||||||||:|||||||||||||:||
Db     361 TASSGYASVQGADEPPLEEHAQSANENLQLVNHEANSVAVDVVPHVDNPTFEEDETPDQE 420

Qy     421 TAVREIKS 428
           ||||||||
Db     421 AAVREIKS 428


RESULT 6
Q9CVG1
ID   Q9CVG1      PRELIMINARY;      PRT;   237 AA.
AC   Q9CVG1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   1300002C13Rik protein (Fragment).
GN   1300002C13RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK008312; BAB25595.1; -.
DR   MGD; MGI:1914139; 1300002C13Rik.
DR   InterPro; IPR003137; PA.
DR   Pfam; PF02225; PA; 1.
FT   NON_TER     237    237
SQ   SEQUENCE   237 AA;  25344 MW;  319E6AA548F7379B CRC64;

  Query Match             53.5%;  Score 1218;  DB 11;  Length 237;
  Best Local Similarity   96.2%;  Pred. No. 3.8e-102;
  Matches  228;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
           |||||| || |||||| :|||||||||||| |||||||||||||||||||||||||||
Db       1 MGPPPGIGVYCRGGCGAARLLAWCFLLALSPHAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60

Qy      61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db      61 TVWELSEEGVYGQDSPLEPVSGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120

Qy     121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
           |||||||||||||| || |||||||||||||||||||||||| ||||||||||||||||||
Db     121 QRGGGCTFADKIHLASERGASGAVIFNFPGTRNEVIPMSHPGAGDIVAIMIGNLKGTKIL 180

Qy     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNA 237
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNA 237


RESULT 7
Q95JW9
ID   Q95JW9      PRELIMINARY;      PRT;   376 AA.
AC   Q95JW9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 42.2 kDa protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AB070058; BAB63003.1; -.
DR   InterPro; IPR001309; ICE_p20.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   PROSITE; PS01121; CASPASE_HIS; UNKNOWN_1.
KW   Hypothetical protein; Zinc-finger.
SQ   SEQUENCE   376 AA;  42248 MW;  B349791C19956841 CRC64;

  Query Match             40.9%;  Score 930;  DB 6;  Length 376;
```

Best Local Similarity 48.6%; Pred. No. 9.4e-76;
Matches 191; Conservative 65; Mismatches 109; Indels 28; Gaps 10;

```
Qy      12 RGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNRTVWELSEEGVY 71
Db      10 RNNTAFSWLIMFGVLWLVSQNC---CRASVVWTAYMNISF---HVG-NHVLSELGETGVF 62

Qy      72 GQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTFADK 131
Db      63 GRSSTLKRVAGVIVPPEGK-IQNACNPNTIFS-----RSKYSETWLALIER-GGCTFTQK 115

Qy     132 IHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTM 191
Db     116 IKVAAEKGASGVIIYNFPGTGNQVFPMFHQAFEDVVVVMIGNLKGTEIFHLIKKGVLITA 175

Qy     192 VIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADA 251
Db     176 MVEVGRKHIIWMNHY-----LVSFVIVTTATLAYFIFYHIHRLCLARIQNRRWQRLTTDL 230

Qy     252 KKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKH 311
Db     231 QNAFGQLQLRVVKEGDEEINPNGDSCVICFELYKPNDIVRILTCKHFFHKNCIDPWILSH 290

Qy     312 KTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQG 371
Db     291 GTCPICKCDILKVLGFQVDVENGTEPLQVLMSSELCETLSPSEEETNNE------VSPAG 344

Qy     372 TDEPPLEEHVQSTNESLQLVNHEANSVAVDVIP 404
Db     345 TSDKVI--HVEE-NPTSQNNDSQPHSVVEDVHP 374
```

RESULT 8
Q95K04
ID Q95K04 PRELIMINARY; PRT; 376 AA.
AC Q95K04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 42.2 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL; AB070023; BAB62968.1; -.
DR InterPro; IPR001309; ICE_p20.
DR InterPro; IPR003137; PA.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PROSITE; PS01121; CASPASE_HIS; UNKNOWN_1.
KW Hypothetical protein; Zinc-finger.
SQ SEQUENCE 376 AA; 42238 MW; D44585F36F4E0C6E CRC64;

Query Match 40.7%; Score 927; DB 6; Length 376;
Best Local Similarity 48.6%; Pred. No. 1.8e-75;
Matches 191; Conservative 65; Mismatches 109; Indels 28; Gaps 10;

```
Qy      12 RGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNRTVWELSEEGVY 71
Db      10 RNNTAFSWLIMFGVLWLVSQNC---CRASVVWTAYMNISF---HVG-NHVLSELGETGVF 62

Qy      72 GQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTFADK 131
Db      63 GRSSTLKRVAGVIVPPEGK-IQNACNPNTIFS-----RSKYSETWLALIER-GGCTFTQK 115

Qy     132 IHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTM 191
Db     116 IKVAAEKGASGVIIYNFPGTGNQVFPMFHQAFEDVVVVMIGNLKGTEIFHLIKKGVLITA 175

Qy     192 VIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADA 251
Db     176 MVEVGRKHIIWMNHY-----LVSFVIVTTATLAYFIFYHIHRLCLARIQNRRWQRLTTDL 230

Qy     252 KKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKH 311
Db     231 QNAFGQLQLRVVKEGDEEINPNGDSCVICFEHYKPNDIVRILTCKHFFHKNCIDPWILSH 290

Qy     312 KTCPMCKCDILKALGIEVDVEDGSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQG 371
Db     291 GTCPICKCDILKVLGIQVDVENGTEPLQVLMSSELCETLSPSEEETNNE------VSPAG 344

Qy     372 TDEPPLEEHVQSTNESLQLVNHEANSVAVDVIP 404
Db     345 TSDKVI--HVEE-NPTSQNNDSQPHSVVEDVHP 374
```

RESULT 9
Q8WVZ7
ID Q8WVZ7 PRELIMINARY; PRT; 376 AA.
AC Q8WVZ7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 42.3 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC022038; AAH22038.1; -.
DR InterPro; IPR003137; PA.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 42294 MW; 7C0446D1FB0981F1 CRC64;

Query Match 40.0%; Score 910.5; DB 4; Length 376;
Best Local Similarity 50.0%; Pred. No. 5.5e-74;
Matches 184; Conservative 58; Mismatches 97; Indels 29; Gaps 9;

```
Qy      39 AEAVWTAYLNVSWRVPHTGVNRTVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNP 98
Db      34 ASVVWMAYMNISF---HVG-NHVLSELGETGVFGRSSTLKRVAGVIVPPEGK-IQNACNP 88

Qy      99 HTNFTVPTVWGSTVQVSWLALIQRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPM 158
Db      89 NTIFS-----RSKYSETWLALIER-GGCTFTQKIKVATEKGASGVIIYNVPGTGNQVFPM 142

Qy     159 SHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFII 218
Db     143 FHQAFEDVVVVMIGNLKGTEIFHLIKKGVLITAVVEVGRKHIIWMNHY-----LVSFVIV 197

Qy     219 TAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCA 278
Db     198 TTATLAYFIFYHIHRLCLARIQNRRWQRLTTDLQNTFGQLQLRVVKEGDEEINPNGDSCV 257

Qy     279 VCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIEVDVEDGSVSL 338
```

```
Db    258 ICFERYKPNDIVRILTCKHFFHKNCIDPWILPHGTCPICKCDILKVLGIQVVVENGTEPL 317

Qy    339 QVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPL--EEHVQSTNESLQLVNHEAN 396
          || :|||:  : |  ||:  :|     |  || :  :  ||:  | |  :|     :
Db    318 QVLMSNELPETLSPSEEETNNE------VSPAGTSDKVIHVEENPTSQNNDIQ-----PH 366

Qy    397 SVAVDVIP 404
          ||  || |
Db    367 SVVEDVHP 374


RESULT 10
Q9D3V4
ID   Q9D3V4      PRELIMINARY;  PRT;   316 AA.
AC   Q9D3V4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   4933432M07Rik protein.
GN   4933432M07RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK017027; BAB30557.1; -.
DR   MGD; MGI:1918550; 4933432M07Rik.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
KW   Zinc-finger.
SQ   SEQUENCE   316 AA;  35535 MW;  153156D58A375991 CRC64;

  Query Match             36.4%;  Score 829;  DB 11;  Length 316;
  Best Local Similarity   51.4%;  Pred. No. 1e-66;
  Matches  166;  Conservative   51;  Mismatches   82;  Indels   24;  Gaps    7;

Qy      1 MGPPPGAGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNR 60
          :|| |   |      |   |    |||     : |    :|:|||:||::::|     ||
Db     17 LGPTP----SVHRSVSFWLLRLSVFLLLSLRDSKG----KAIWTAHLNITFQVG----NR 64

Qy     61 TVWELSEEGVYGQDSPLEPVAGVLVPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALI 120
           : || | ||:|  |||| |:| :| |:|    |||:| |||: |         :|||||
Db     65 IISELGESGVFGNHSPLERVSGAVVLPEGWNQ-NACSPLTNFSRPDQTD-----TWLALI 118

Qy    121 QRGGGCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKIL 180
          :| |||||  ||:|| |: |:| :|:|:||| |:| |||| |  :|||:|||||| ::|
Db    119 ER-GGCTFTHKINLAAEKGANGVIIYNYPGTGNKVFPMSHQGTENIVAVMIGNLKGMELL 177

Qy    181 QSIQRGIQVTMVIEVGKKHGPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQ 240
            ||:|: ||::|||| | ||::||      :| |   |||| |   | | | | : :
Db    178 HLIQQGVYVTIIIEVGRMHMPWLSHY-----VMSLFTFLAATVTYLFLYCAWRPRVSNSS 232

Qy    241 SRKQRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFH 300
          :|:||||||| |||||:|||| |:  |||: |: ||| || ::||  |::|||| | ||
Db    233 TRRQRQLKADVKKAIGQLQLRVLQDEDKELDPNEDSCVVCFDMYKAQDVIRILTCKHFFH 292

Qy    301 KTCVDPWLLKHKTCPMCKCDILK 323
          |||:||||| |:||||||||||
Db    293 KTCIDPWLLAHRTCPMCKCDILK 315


RESULT 11
Q9H5E4
ID   Q9H5E4      PRELIMINARY;      PRT;   156 AA.
AC   Q9H5E4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE-  CDNA: FLJ23516 fis, clone LNG04848.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA   Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA   Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA   Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK027169; BAB15682.1; -.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
KW   Zinc-finger.
SQ   SEQUENCE   156 AA;  17263 MW;  526F427554DD4FB9 CRC64;

  Query Match             31.4%;  Score 714;  DB 4;  Length 156;
  Best Local Similarity   89.7%;  Pred. No. 1e-56;
  Matches  139;  Conservative    5;  Mismatches    3;  Indels    8;  Gaps    3;

Qy    277 CAV-CIELYKPNDLV--RILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKALGIEVDVED 333
          ||: ||    | ::   |||||||||||||||||:|:||||||||||||||||||||
Db      7 CALNCI-----NQMIWYSILTCNHIFHKTCVDPWLLEHRTCPMCKCDILKALGIEVDVED 61

Qy    334 GSVSLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVNH 393
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 GSASLQVPVSNEISNSASSHEEDNRSETASSGYASVQGTDEPPLEEHVQSTNESLQLVNH 121

Qy    394 EANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 428
          ||||||||||||||||||||||||||||||||||||
Db    122 EANSVAVDVIPHVDNPTFEEDETPNQETAVREIKS 156


RESULT 12
Q8WUU3
ID   Q8WUU3      PRELIMINARY;      PRT;   426 AA.
AC   Q8WUU3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 45.7 kDa protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC019355; AAH19355.1; -.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   426 AA;  45687 MW;  EF2B697A41368451 CRC64;

  Query Match             28.6%;  Score 650.5;  DB 4;  Length 426;
  Best Local Similarity   37.6%;  Pred. No. 2.4e-50;
  Matches  160;  Conservative   65;  Mismatches  149;  Indels   51;  Gaps   13;

Qy       7 AGVSCRGGCGFSRLLAWCFLLALSPQAPGSRGAEAVW-TAYLNVSWRVPHTGVNRTVWEL 65
           | |  ||      :||   ||||:   ||:||    | :| :|: :  | |  | ||| :
Db      34 ASVGARG------VLALA-LLALALCVPGARGRALEWFSAVVNIEYVDPQT--NLTVWSV 84

Qy      66 SEEGVYGQDSPLEPVAGVLVPPDGPGA-LNACNPHTNFTVPTVWGSTVQVSWLALIQRGG 124
           || | :|  || |   |::  |  ||  |  | | | ||    |      |:|| | |
Db      85 SESGRFGDSSPKEGAHGLVGVPWAPGGDLEGCAPDTRFFVPEP-GGRGAAPWVALVAR-G 142

Qy     125 GCTFADKIHLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQ 184
           |||| ||: :|  | ||  |::|     |  :|||| |  :|| |||   || :||: :|
Db     143 GCTFKDKVLVAARRNASAVVLYNEERYGNITLPMSHAGTGNIVVIMISYPKGREILELVQ 202

Qy     185 RGIQVTMVIEVGKKH-GPWVNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRK 243
           :|| ||| | || :|   :::  |: ||:::|  :   :: : |||  :|     :|
Db     203 KGIPVTMTIGVGTRHVQEFISGQSVVFVAIAFITMMIISLAWLIFYYIQRFLYTGSQIGS 262

Qy     244 QRQLKADAKKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTC 303
           |   | : || ||:| | |:| |:| |  | ::|||||| :|  |::||| | ||||: |
Db     263 QSHRK-ETKKVIGQLLLHTVKHGEKGIDVDAENCAVCIENFKVKDIIRILPCKHIFHRIC 321

Qy     304 VDPWLLKHKTCPMCKCDILKALGI---EVDVEDGSVSLQVPVSNEISNSASSHEEDNRSE 360
           :||||| |:|||||| |::||||     ||::      |  :  :| : :  :|: |:
Db     322 IDPWLLDHRTCPMCKLDVIKALGYWGEPGDVQEMPAPESPPGRDPAANLSLALPDDDGSD 381

Qy     361 TASSGYASVQGTDEPPLEEHVQSTNESLQLVNHEANSVAVDVIPHVDNPTFEEDETPNQE 420
           :|   ||       | |   |                          :|:|: |   |
Db     382 ESSPPSAS-------PAESEPQC----------------------DPSFKGDAGEN-- 408

Qy     421 TAVRE 425
           ||: |
Db     409 TALLE 413


RESULT 13
Q8VEM1
ID   Q8VEM1      PRELIMINARY;      PRT;   419 AA.
AC   Q8VEM1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   G1-related zinc finger protein.
GN   G1RZFP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC018199; AAH18199.1; -.
DR   MGD; MGI:1891717; G1rzfp.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   419 AA;  46375 MW;  4E699A66D658CBF2 CRC64;

  Query Match             26.2%;  Score 595.5;  DB 11;  Length 419;
  Best Local Similarity   42.4%;  Pred. No. 2.3e-45;
  Matches  134;  Conservative   49;  Mismatches  120;  Indels   13;  Gaps    7;

Qy      16 GFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNRTVWELSEEGVYGQDS 75
           | :|| |   |            ::  :|| :|: : |  |    |     : | || ||
Db       8 GPARLAALALLTCSLWPTRADNASQEYYTALINVTVQEPGRGTPLTF--RIDRGRYGLDS 65

Qy      76 PLEPVAGVLVPP---DGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTFADKI 132
           |   | | :: |     |       |:| | | ||      ::  |:||:|| | ||| :||
Db      66 PKAEVRGQVLAPLPIHGVADHLGCDPQTRFFVP----PNIK-QWIALLQR-GNCTFKEKI 119

Qy     133 HLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTMV 192
             |    |   ||:|   :: | : |:|||  ||:|:||  |:|  ||  ::: | | |
Db     120 SRAAFHNAVAVVIYN-NKSKEEPVTMTHPGTGDIIAVMITELRGKDILSYLEKNISVQMT 178

Qy     193 IEVGKKHGPW-VNHYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADA 251
           | || :   |   :  |: |||:|| ::   :  : |||  :::|   |: | ||:|   |
Db     179 IAVGTRMPPKNFSRGSLVFVSISFIVLMIISSAWLIFYFIQKIRYTNARDRNQRRLGDAA 238

Qy     252 KKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKH 311
           |||| :|  ||:|:|||| || | |||||| || ||:||:| | |:|||:|||||| :|
Db     239 KKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVRVLPCKHVFHKSCVDPWLSEH 298

Qy     312 KTCPMCKCDILKALGI 327
           ||||||  :|||||||
Db     299 CTCPMCKLNILKALGI 314


RESULT 14
Q9QZQ6
ID   Q9QZQ6      PRELIMINARY;      PRT;   419 AA.
AC   Q9QZQ6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   G1-related zinc finger protein.
GN   G1RZFP OR G1RP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20267845; PubMed=10806348;
RA   Baker S.J., Reddy E.P.;
RT   "Cloning of murine G1RP, a novel gene related to Drosophila
RT   melanogaster g1.";
RL   Gene 248:33-40(2000).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF171875; AAF05310.1; -.
DR   MGD; MGI:1891717; G1rzfp.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW   Zinc-finger.
SQ   SEQUENCE   419 AA;  46403 MW;  F3DF4118EEB61A01 CRC64;

  Query Match             26.0%;  Score 592.5;  DB 11;  Length 419;
```

Best Local Similarity 42.1%; Pred. No. 4.2e-45;
Matches 133; Conservative 50; Mismatches 120; Indels 13; Gaps 7;

```
Qy      16 GFSRLLAWCFLLALSPQAPGSRGAEAVWTAYLNVSWRVPHTGVNRTVWELSEEGVYGQDS 75
Db       8 GPARLAAALALLTCSLWPTRADNASQEYYTALINVTVQEPGRGTPLTF--RIDRGRYGLDS 65

Qy      76 PLEPVAGVLVPP---DGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTFADKI 132
Db      66 PKAEVRGQVLAPLPIHGVADHLGCDPQTRFFVP----PNIK-QWIALLQR-GNCTFKEKI 119

Qy     133 HLAYERWASGAVIFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTMV 192
Db     120 SRAAFHNAVAVVIYN-NKSKEEPVTMTHPGTGDIIAVMITELRGKDILSYLEKNISVQMT 178

Qy     193 IEVGKKHGPW-VNHYSIFFVSVSFPIITAATVGYFIFYSARRLRNARAQSRKQRQLKADA 251
Db     179 IAVGTRMPPKNFSRGSLVFVSISFIVLMIISSAWLIFYFIQKIRYTNARDRNQRRLGDAA 238

Qy     252 KKAIGRLQLRTLKQGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKH 311
Db     239 KKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVRVLPCKHVFHKSCVDPWLSEH 298

Qy     312 KTCPMCKCDILKALGI 327
Db     299 CTCPMCKLNILRALGI 314
```

RESULT 15
Q9ULK6
ID Q9ULK6 PRELIMINARY; PRT; 462 AA.
AC Q9ULK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1214 protein (Fragment).
GN KIAA1214.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirosawa M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL; AB033040; BAA86528.1; -.
DR InterPro; IPR003137; PA.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
KW Zinc-finger.
FT NON_TER 1 1
SQ SEQUENCE 462 AA; 49883 MW; D1279DE127C9772D CRC64;

Query Match 25.8%; Score 586; DB 4; Length 462;
Best Local Similarity 35.6%; Pred. No. 1.9e-44;
Matches 156; Conservative 46; Mismatches 124; Indels 112; Gaps 11;

```
Qy       6 GAGVSCRGGCGFSR----------------------LLAWCF--LLALSPQAPGSRGAE 40
Db      47 GGGTGRREGAGAGRGPRRIKMAMSLIQACCSLALSTWLLSFCFVHLLCLDFTVAEK---E 103

Qy      41 AVWTAYLNVSWRVP---------------HTGVNRTVWELSEEGVYGQDSPLEPVAGVL 84
Db     104 EWYTAFVNITYAEPAPDPGAGAAGGGGAELHT-------EKTECGRYGEHSPKQDARGEV 156

Qy      85 VPPDGPGALNACNPHTNFTVPTVWGSTVQVSWLALIQRGGGCTFADKIHLAYERWASGAV 144
Db     157 VMASSAHDRLACDPNTKFAAP-----TRGKNWIALIPK-GNCTYRDKIRNAFLQNASAVV 210

Qy     145 IFNFPGTRNEVIPMSHPGAVDIVAIMIGNLKGTKILQSIQRGIQVTMVIEVGKKHGPWVN 204
Db     211 IFNVGSNTNETITMPHAGVEDIVAIMIPEPKGKEIVSLLERNITVTMYITIGTRN----- 265

Qy     205 HYSIFFVSVSFFIITAATVGYFIFYSARRLRNARAQSRKQRQLKADAKKAIGRLQLRTLK 264
Db     266 ----------------------------------LQKYRRLGDAAKKAISKLQIRTIK 289

Qy     265 QGDKEIGPDGDSCAVCIELYKPNDLVRILTCNHIFHKTCVDPWLLKHKTCPMCKCDILKA 324
Db     290 KGDKETESDFDNCAVCIEGYKPNDVVRILPCRHLFHKSCVDPWLLDHRTCPMCKMNILKA 349

Qy     325 LGIEVDVE-------DGSVSLQVPVSNEISNSASSHEEDNRSETASSGYAS------VQG 371
Db     350 LGIPPNADCMDDLPTDFEGSLGGPPTNQIT-GASDTTVNESSVTLDPAVRTVGALQVVQD 408

Qy     372 TDEPPLEEHVQSTNESLQ 389
Db     409 TDPIPQEGDVIFTTNSEQ 426
```

Search completed: April 22, 2003, 08:26:58
Job time : 38 secs